~~libertad~~ artificial

Discursos, redes y pluralidad.
Impactos diferenciados en la moderación de contenidos en plataformas digitales

>:(<3

Francisco Chan Chan
Adriana A. Figueroa Muñoz Ledo
Jesús Eulises González Mejía

Coordinadores
Priscilla Ruiz y Pablo Pruneda Gross

A partir de esfuerzos compartidos de
Artículo 19, Oficina para México y Centroamérica
Línea de Investigación en Derecho e Inteligencia Artificial, IIJ UNAM
Laboratorio Nacional Diversidades, IIJ UNAM
Cultivando Género, A.C.
Colectivo por la Protección de Todas las Familias en Yucatán

Diseño de identidad y editorial
Fósforo - fsfr.mx

libertad artificial

Discursos, redes y pluralidad.
Impactos diferenciados en la moderación de contenidos en plataformas digitales

Contenido

Presentación

Hoy en día, nuestra realidad está mediada por lo digital como en ninguna otra época de la historia[1]. Internet y sus aplicaciones para las comunicaciones son ahora un medio clave que crea nuevas formas de comunidad. Desde el punto de vista de los derechos humanos, Internet se ha convertido en un medio importante para ejercer la libertad de expresión y el correlativo acceso a la información[2]. Lo anterior se articula coherentemente con los principios que caracterizan al Estado moderno, es decir, con los derechos a la información y a la libertad de expresión, mismos que, junto con el derecho a la libertad de asociación y derecho a la reunión se consideran fundamentales para la naturaleza participativa de la democracia[3]. Esto otorga a las sociedades contemporáneas herramientas de diálogo para la toma de decisiones por la vía democrática, al tiempo que su relevancia se profundiza en otras libertades y derechos.

01 Asociación de Internet MX. (Mayo 2022). 18° Estudio sobre los hábitos de personas usuarias de Internet en México. Disponible en: https://www.asociaciondeinternet.mx/estudios/asociacion

02 Libertad de expresión e Internet, Relatoría Especial para la Libertad de Expresión, Comisión Interamericana de Derechos Humanos, OEA, Open Society Foundations, diciembre 2013, p. 6. https://www.oas.org/es/cidh/expresion/docs/informes/2014_04_08_Internet_WEB.pdf

03 Revilla, M. Participación política: lo individual y lo colectivo en el juego democrático. En Benedicto, J. y Morán, M. (eds.). Sociedad y política. Temas de sociología política. Alianza editorial. Segunda reimpresión. Madrid; 2017. Véase https://www.ohchr.org/en/calls-for-input/2020/call-comment-no-37-article-21-international-covenant-civil-and-political

En este contexto se han producido diversos cambios. En el campo de los derechos sociales, los relativos a educación y salud[4] han modificado la forma en que son ejercidos. Todos fuimos testigos de la mudanza imprevista a las clases a distancia y el trabajo remoto y de regreso en una permanente incertidumbre durante la pandemia. Sus efectos se encuentran aún en estudio, pero existen datos evidentes de rezago educativo y un aumento de las múltiples brechas que atraviesan a la juventud y niñez mexicana.[5] En la cuestión de servicios, tanto los públicos como los privados se han apoyado en medios tradicionales e Internet para superar el distanciamiento social de la pandemia. Respecto a la relación entre la ciudadanía y las instituciones, en México de 2019 a 2021 se reportó una escalada de 22 puntos porcentuales en la población de 18 años y más, que ha interactuado con el gobierno a través de Internet, pasando de 32.4 a 54.5 por ciento.[6] Por su parte, el acceso a la justicia y los tribunales digitales se han convertido en una obligación constitucional[7] y algunos Estados están avanzando sustancialmente en mejorar sus sistemas de justicia incorporando las nuevas tecnologías.[8]

En ese sentido, Internet ya no es sólo un medio para difundir y publicar información, es también una herramienta necesaria para el ejercicio de los derechos humanos. Cada vez más, la red se convierte en una parte integral de la vida laboral, familiar y social, de la educación, de la expresión y de la construcción de nuestra identidad. En particular, las redes sociodigitales se han convertido en foros virtuales donde muchas personas discuten sobre todos los temas con variados márgenes de civilidad.[9]

Como cualquier otro espacio social humano, Internet reproduce dinámicas y prácticas sociales del espacio *offline*. La violencia, la desigualdad y la injusticia se reflejan en las vidas digitales, así como las brechas económicas, educativas y sanitarias de las desigualdades por razones de género, raciales, etarias, de clase social, en el entorno lo rural y urbano, que crecen y se profundizan en nuestras comunidades.[10]

Al convertirse en un lugar común para millones de seres humanos en el mundo, en las redes sociodigitales se han desarrollado algunas prácticas y en el caso de México una reforma a los marcos jurídicos para contener la violencia y reducir la desigualdad,[11] para lo cual, las plataformas digitales han creado mecanismos

04 Asociación de Internet MX. (Abril 2021). 1er estudio sobre los hábitos de Médicos de Internet en México. Estadística digital. Central Media Agencia Digital. Close-up. GSK. PLM. Smart scale. Disponible en: https://irp.cdn-website.com/81280eda/files/uploaded/Estudio%20sobre%20los%20Ha%CC%81bitos%20de%20los%20Me%CC%81dicos%20en%20Internet%20en%20Me%CC%81xico%20AIMX%202021%20versio%CC%81n%20pu%CC%81blica.pdf

05 "El rezago educativo es inevitable porque las familias tuvieron muchos problemas (salud física, mental, economía o conectividad) que los obligaron a no preocuparse por las clases virtuales de sus hijos durante casi un año y medio. En México y a nivel mundial, algunos estudiantes, gracias a su economía familiar, tuvieron tutores virtuales, clases en grupos reducidos o clases privadas presenciales en grupos reducidos (pandemic pods), lo que aumentó la brecha socioeducativa porque quienes no tenían los medios económicos se quedaron con una educación básica. Por diferentes razones, una buena cantidad de estudiantes de secundaria que antes tenían buenas notas bajaron sus promedios y aprendizajes por lo que se verán afectados en sus posibilidades de ingresar a la preparatoria. El rezago educativo será una de las consecuencias más duras en México debido al manejo de la pandemia […]". Gallegos de Dios, O. (2022). Ausentismo, deserción escolar y rezago educativo en secundarias públicas en México durante la pandemia del Covid-19. Sincronía, n.o 81 (Enero-Junio): 725-45.

06 Instituto Nacional de Estadística y Geografía . Encuesta Nacional de Calidad e Impacto Gubernamental 2021. Disponible en: https://www.inegi.org.mx/programas/encig/2021/

07 Diario Oficial de la Federación. Decreto por el que se declara(n) reformadas y adicionadas diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, relativos al Poder Judicial de la Federación. 11 de marzo de 2021. https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_246_11mar21.pdf

08 Escamilla, S. y Pantin, L. (2021). La Justicia Digital En México: El Saldo A Un Año Del Inicio De La Pandemia - México Evalúa. México Evalúa. Disponible en: https://www.mexicoevalua.org/la-justicia-digital-en-mexico-el-saldo-a-un-ano-del-inicio-de-la-pandemia/

09 Relatoría Libertad de Expresión, 2013, obra citada.

10 Padrón, M. (2014). Disponibilidad y acceso a la tecnología como una aproximación para el estudio del fenómeno de acceso a la información y su relación con la pobreza en México, en Luna, Issa, y Universidad Nacional Autónoma de México. 2014. Estudios aplicados sobre la libertad de expresión y el derecho a la información. 1° ed. México D.F.: Universidad Nacional Autónoma de México Instituto de Investigaciones Jurídicas.

11 Cfr. Flew, T., Martin, F. y Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. Journal of Digital Media and Policy, 10(1), pp. 33-50.

de control del contenido considerado por ellas como “nocivo” y “problemático”; dichos son lo que conocemos como *moderación de contenido*. Este tipo de moderación –que puede ser humana o automatizada– constituye el foco de la presente investigación, en la cual analizamos sus posibilidades de aplicación masiva y los efectos diferenciados que, en el campo de la libertad de expresión y el acceso a la información, produce en algunos grupos de personas en situación de vulnerabilidad.

Referirnos a la vulnerabilidad como una condición, implica que no la consideramos como un atributo o característica de las personas, sino como una circunstancia que existe, persiste y se intensifica en razón de contextos de desigualdad social, que puede definirse como “*[…] las asimetrías en la capacidad de apropiación de los recursos y activos productivos (ingresos, bienes, servicios, entre otros) que constituyen o generan bienestar entre distintos grupos sociales*”.[12] La desigualdad es consecuencia de la injusticia social que excluye a ciertos grupos a partir de una serie de obstáculos para la redistribución de recursos (materiales y de otra índole) y el reconocimiento o la dimensión ético y cultural de sus diferencias y particularidades.[13] Hay un gran porcentaje de la riqueza mundial concentrada en pocos grupos y también, la exclusión incluye el reparto inequitativo de poder político, razón por la cual las decisiones que tienen impacto en la mayoría de las personas son tomadas por grupos reducidos.[14] La desigualdad social puede expresarse en todos los ámbitos de la vida de las personas, incluyendo, por ejemplo, su posibilidad de conectividad a Internet y sus recursos disponibles para mantener su presencia en el espacio digital.

En la presente investigación hemos considerado a los siguientes grupos sociales para comprender las situaciones diferenciadas de vulnerabilidad: mujeres, personas LGBTQ+, personas indígenas y afromexicanas, personas con discapacidades, periodistas, personas migrantes y defensoras de derechos humanos. Las violencias de que suelen ser objeto estas poblaciones descansan en formas de violencia estructural, o la negación de los derechos humanos básicos, que de manera sistemática reproducen relaciones desiguales de poder y colocan a ciertos grupos en mayor vulnerabilidad que otros, restringiendo así sus oportunidades para satisfacer necesidades y para tener una vida plena.[15] La violencia estructural se sustenta en formas de violencia cultural y violencia directa.[16] Ejemplos de estas violencias son el sexismo, el racismo y el capacitismo.[17]

12 CEPAL, La matriz de la desigualdad social en América Latina, LC/G.2690(MDS.1/2), octubre, 2016, p. 18.

13 Fraser, N. (1997). ¿De la redistribución al reconocimiento? Dilemas en torno a la justicia en una época “postsocialista.” In Iustitia Interrupta (pp. 17–54). Santafé de Bogotá: Siglo del Hombre Editores y Universidad de los Andes, Facultad de Derecho.

14 CEPAL, La matriz de la desigualdad social en América Latina, LC/G.2690(MDS.1/2), octubre, 2016, p. 18.

15 Organización de las Naciones Unidas, Informe del Relator Especial sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental, A/HRC/41/34, 12 de abril de 2019, párr. 86.

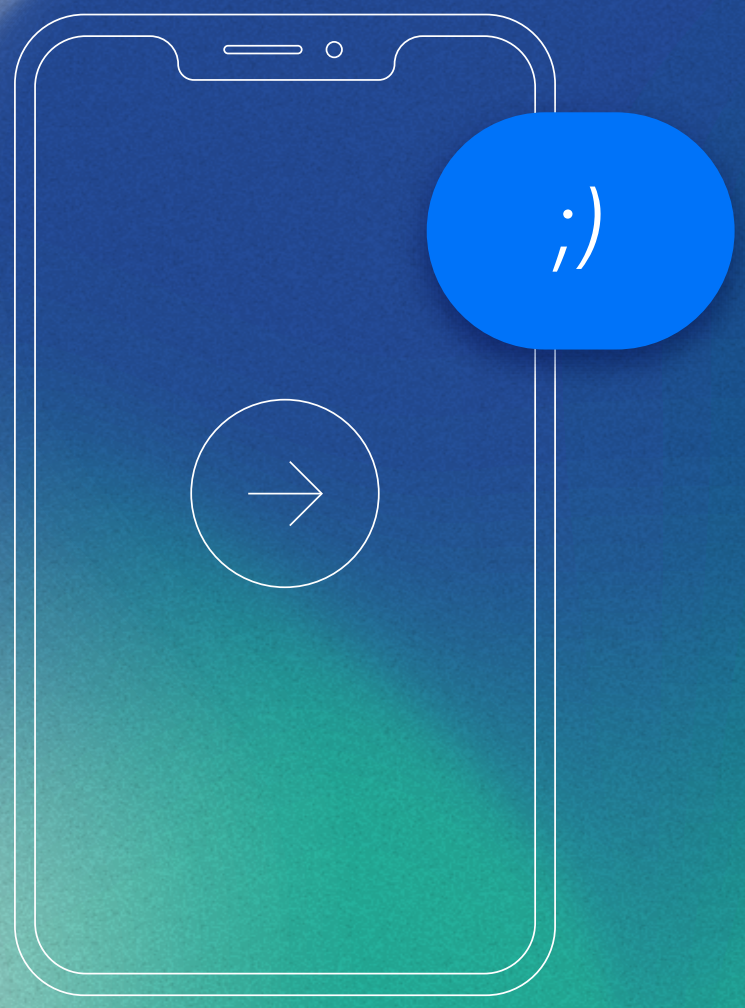
16 Galtung, J. (1990). Cultural Violence. Journal of Peace Research, 27(3), 291–305.

17 “Con “capacitismo” se denomina la expresión de una red de creencias, prácticas y representaciones que establecen una estratificación funcional con la finalidad de institucionalizar la capacidad como la norma; ciertamente esa normatividad es la que forja la estandarización de los cuerpos, proyectando ciertos cuerpos como perfectos, eficientes y válidos ejemplos de la humanidad, mientras que, remarca la existencia de cuerpos que desbordan la condición humana por ser deformes, feos, inválidos y deficientes; cuerpos que están expuestos al contacto con los otros, pero también a la violencia que procura la permanencia de lo normal.” Maldonado, J. (2017, marzo 18). Repensar la práctica del cuidado en el contexto del síndrome de Down. Debate Feminista, 53. https://doi.org/https://doi.org/10.1016/j.df.2017.01.002

Los objetivos de la investigación son, por un lado, explorar el derecho a la libertad de expresión e información en plataformas digitales a través de las perspectivas y voces de personas que enfrentan exclusión y discriminación, y por otro lado, conocer el impacto de la inteligencia artificial, la desinformación y la moderación de contenidos en el ejercicio cotidiano de sus derechos humanos, y el creciente papel y poder de las plataformas digitales y otros actores. Todo ello mediante un análisis de la aplicación de las normas comunitarias[18] en México y su impacto diferenciado en diversas comunidades.

El presente informe se divide en cinco apartados. La *primera parte*, presenta algunas ideas clave para comprender en qué consiste la moderación de contenidos, quién la realiza y bajo qué contextos. Asimismo, se presenta información sobre cuándo ocurre y cómo se lleva a cabo. La *segunda sección*, aborda el derecho a la libertad de expresión y su relación con la moderación de contenidos, especialmente para analizar el tema de censura, autocensura y los efectos no deseados de la moderación automatizada. En un *tercer apartado*, se presentan algunos indicadores básicos sobre los grupos en condición de riesgo y vulnerabilidad que se abordaron en esta investigación; y se presentan algunos argumentos sobre las formas de exclusión y discriminación dentro y fuera de las redes socio digitales. En el *cuarto apartado*, se muestran los hallazgos principales de la investigación, integrando el análisis de la encuesta aplicada, y las entrevistas y los talleres realizados. Se analizan los temas de moderación automatizada y sus efectos, la libertad de expresión y las brechas de acceso y ejercicio de derechos en el ámbito digital. Finalmente, en el *quinto apartado*, se presentan las conclusiones y las recomendaciones puntuales para mejorar las condiciones de moderación de contenido, libertad de expresión y pluralidad de participación en las redes sociodigitales.

18 Nos referimos a los instrumentos jurídicos “consensuales” en los que las Plataformas establecen los contenidos no permitidos y sus sanciones.



Estrategia metodológica del proyecto

El presente trabajo es resultado de un estudio exploratorio con enfoque participativo. Dado que el proyecto abordó una realidad diversa y compleja como es la moderación de contenidos y sus efectos en distintos grupos sociales, se dio prioridad a la participación de distintas voces en el contexto mexicano para ampliar el panorama respecto a la libertad de expresión y acceso a la información. Los resultados que se presentan en el presente documento no pretenden representatividad estadística, es decir, no hay interés en extrapolar los resultados a poblaciones más amplias. En su lugar, buscamos entablar un diálogo a detalle para explorar experiencias diversas respecto a moderación de contenido, libertad de expresión y acceso a la información. Lo anterior vinculado con el propósito de generar puentes de entendimiento y proponer soluciones articuladas desde distintos puntos de vista.

Como parte de nuestro esfuerzo por dar espacio a la diversidad de perspectivas y experiencias, la planificación metodológica incluyó diversas actividades con enfoque participativo e interdisciplinario, desde el diseño de instrumentos de recolección de información, la selección de conceptos y categorías de análisis, y la toma de decisiones analíticas y operativas. Asimismo, sostenemos la convicción de que el conocimiento se construye a través de alianzas colaborativas entre la academia, organizaciones de la sociedad civil y otros colectivos y de la discusión permanente entre las distintas comunidades. Desde este enfoque participativo, incorporamos la diversidad de voces como una forma de colaboración cercana con todas las personas que directa e indirectamente participaron a lo largo de todo el proceso de investigación. Se procuró en todo momento fortalecer los esfuerzos realizados para fortalecer y visibilizar ante la gran red de redes de agentes de cambio que buscan transformar la segregación y la exclusión en acciones colectivas de resistencia.

Priorizamos también la construcción de categorías analíticas robustas para comprender a los grupos sociales y a los sujetos en su complejidad. Al respecto, no podemos hablar de las mujeres, de las personas de la comunidad LGBTQ+, de personas indígenas y afromexicanas, personas migrantes, periodistas y activistas como sujetos homogéneos. Subrayamos este punto ya que con la diversidad de voces presentadas en este informe no pretendemos hablar de todas las experiencias, ni tampoco consideramos que nuestras(os) interlocutoras(es) representan la totalidad de un colectivo; es decir, reconocemos que las colectividades no son homogéneas, estáticas ni monolíticas. Por el contrario, aceptamos que cuando hablamos de, por ejemplo “mujeres”, es imposible dar cuenta de las realidades de todas En ese sentido, los resultados de este estudio derivan de un diálogo acotado, pero relevante para insistir en la defensa de los derechos a la libertad de expresión y acceso a la información para poblaciones en condición de vulnerabilidad.

Ligado a lo anterior, consideramos que el ejercicio de los derechos humanos no es inmutable ni predeterminado, y sus efectos deben ser analizados en contexto. Por ello, abordamos a las comunidades de interés de este estudio desde una perspectiva interseccional, para analizar de qué manera categorías como género, clase social y etnicidad, entre otras, se articulan con la forma en que opera la moderación de contenidos. Además, rescatamos la capacidad de agencia de las y los sujetos y comunidades de estudio dentro del ámbito de lo digital.

Al tratarse de un estudio de corte cualitativo, el proyecto exploró las experiencias y significados que las personas construyen en torno a su interacción con la moderación de contenidos en el ámbito digital. Este conocimiento abre la posibilidad de reinterpretar los datos duros que ya existen y trazar un panorama que dé pie a investigaciones más amplias, pero siempre desde el conocimiento de las experiencias de los diversos grupos con los que se trabajó, enmarcadas en condiciones de desigualdad más profundas que exceden y anteceden a las y los individuos.

Desde estas bases, hemos construido un conjunto de recomendaciones para prevenir y minimizar los efectos negativos que supone Internet para las personas en situación de vulnerabilidad, desde los diferentes actores que participan en las plataformas digitales. Para alcanzar este cometido: **a)** mediante entrevistas, aplicación de encuesta y diálogos dentro de los talleres, recabamos experiencias

en torno a moderación de contenido de personas pertenecientes a diversos grupos que pueden ser considerados en condición de vulnerabilidad: mujeres, personas de la comunidad LGBTQ+, personas con alguna discapacidad, personas indígenas y afromexicanas, personas migrantes y personas activistas para con alguna de estas poblaciones; **b)** de la revisión documental, entrevistas a informantes clave e interacción con las personas participantes de los talleres identificamos cuáles han sido las normas, políticas o criterios de las plataformas aplicables y los principales retos y obstáculos de las que comunidades en situación de vulnerabilidad (enunciadas en el punto anterior) se enfrentan en el ejercicio de su libertad de expresión y acceso a la información dentro del espacio digital; y **c)** tanto en el grupo de investigación como en los talleres discutimos en torno temas como el funcionamiento de los algoritmos de moderación de contenido, y principales problemas y buenas prácticas para el control de la desinformación.

El diseño de los métodos e instrumentos de recolección de información se realizó a partir de las siguientes preguntas eje:

01 **¿Cómo se ejerce el derecho a la libertad de expresión e información en plataformas digitales?**

02 **¿Qué formas y prácticas de exclusión y discriminación enfrentan diversos grupos para ejercer su derecho a la libertad de expresión e información en plataformas digitales?**

03 **¿Cómo se articulan la inteligencia artificial, la desinformación y la moderación de contenidos en la producción de dichas prácticas de exclusión y discriminación?**

04 **¿Cómo experimentan y enfrentan diversos grupos los efectos diferenciados de la moderación de contenidos?**

01 Técnicas e instrumentos

Para alcanzar los objetivos planteados, se diseñaron diversas técnicas e instrumentos en correspondencia con los siguientes ejes:

Eje de investigación

Documental. Por una parte, se realizó una revisión de literatura reciente sobre moderación de contenido en plataformas digitales para contar con un panorama general de la discusión actual. Por otro lado, se hizo una labor de recopilación y análisis documental de normas jurídicas, instituciones, doctrina y conceptos relacionados con: i) la moderación de contenido en las plataformas digitales y ii) grupos de personas en situación de vulnerabilidad. Este apartado se ha trabajado desde dos núcleos de investigación del Instituto de Investigaciones Jurídicas de la UNAM: i) la Línea de Investigación en Derecho e Inteligencia Artificial y ii) el Laboratorio Nacional Diversidades.

Entrevistas semiestructuradas. Se realizaron un total de 11 entrevistas de visión experta con informantes clave con el propósito de rescatar sus opiniones y experiencias en torno a la moderación de contenido en redes sociodigitales en articulación con las categorías de análisis (género y diversidad sexo-genérica, condición migrante, etnicidad y discapacidad). El criterio de selección de las personas entrevistadas fue que tuvieran una autoadscripción a alguno de los grupos mencionados previamente, que realizaran algún tipo de activismo en favor de dichos grupos y/o que contaran con conocimiento experto sobre Internet e inteligencia artificial. Las personas entrevistadas fueron contactadas por diversos medios en los que se les hicieron explícitos los propósitos y alcances de su participación y de la investigación. Las entrevistas se realizaron entre febrero y abril de 2022 a través de la plataforma Zoom por integrantes del equipo de trabajo tomando como referencia la guía de entrevista previamente elaborada (véase Anexo 2).

Encuesta. Este instrumento se aplicó a través de la plataforma Limesurvey a las personas asistentes a los talleres (véase Anexo 3).

La información recabada por medio de la encuesta se empleó para: a) elaborar un perfil sociodemográfico general de las personas asistentes a los talleres (8 ítems), b) conocer sus prácticas principales en redes sociodigitales (8 ítems) y c) conocer sus experiencias generales con el ejercicio del derecho a la libertad de expresión en espacios digitales (24 ítems). Respecto a los puntos b y c, esta información sirvió además para retroalimentar y ajustar los contenidos presentados en los talleres. Se aplicaron un total de 52 encuestas entre febrero y abril de 2022, priorizando la mayor diversidad posible de personas participantes.

Eje de intervención-investigación

Talleres. Entre febrero y mayo de 2022 se realizaron tres talleres presenciales en la capital de Aguascalientes (22 de febrero), en Mérida, Yucatán (16 marzo) , y Tijuana (31 marzo); además de un taller en línea impartido desde la Ciudad de México el 13 mayo. El objetivo de estos talleres fue explorar y discutir las experiencias sobre la moderación de contenidos y su relación con el derecho a la libertad de expresión e información en redes digitales. Las temáticas específicas que se abordaron fueron: libertad de expresión en México, moderación de contenido, seguridad digital, impacto de las redes sociodigitales en los movimientos sociales actuales, relación entre algoritmos y libertad de expresión, desinformación y discursos de odio. Los contenidos de los talleres se formularon por el equipo de trabajo y se retroalimentaron con información derivada de las primeras encuestas y entrevistas realizadas. Toda la información recopilada sobre las experiencias de las personas participantes fueron sistematizadas y analizadas, lo cual se refleja en las recomendaciones del estudio.

Eje de comunicación-difusión

Campaña de comunicación. En el marco de este proyecto se diseñó y desarrolló una campaña que tuvo por objetivo dar a conocer a públicos interesados el trabajo de investigación-intervención realizado, así como elaborar productos de comunicación creativos en aras de informar y alentar la discusión pública en torno a la moderación de contenido en espacios digitales. Si bien se contó con un equipo de especialistas en comunicación, los contenidos fueron socializados y

discutidos en todo momento en reuniones grupales con todo el equipo de trabajo para recibir retroalimentación y la aprobación de los distintos productos (flyers de difusión, pósters, stickers).

Seminario. Con el objetivo de trasladar las discusiones y resultados del proyecto a un espacio para compartir y socializar la información, se realizó el Seminario “Inteligencia Artificial, remoción de contenido y libertad de expresión en México” los días 20 y 21 de junio en la Ciudad de México. El evento se realizó en formato híbrido, es decir, presencial previo registro de asistentes y vía streaming a través de las redes sociodigitales de Artículo 19. Se contó con la presencia de 15 ponentes con distintos perfiles: activistas, académicas(os), personas servidoras públicas, especialistas en inteligencia artificial y una representante de Twitter en México. El seminario estuvo organizado en cuatro mesas de discusión (dos por día), cuyas temáticas fueron las siguientes: a) mesa 1: Inteligencia artificial y su impacto en la libertad de expresión, b) mesa 2: Desinformación en el entorno digital, c) mesa 3: Discursos estigmatizantes e incitación al odio dentro del entorno digital, y d) mesa 4: Impacto diferenciado en la moderación de contenido.

Twitter spaces. La reciente función de la red sociodigital Twitter, que permite la transmisión en directo, fue empleada como espacio para la difusión del proyecto. En un primer momento (19 de mayo) se realizó una discusión general sobre la moderación de contenido y su relación con el ejercicio de derechos; posteriormente (21 de junio) se llevó a cabo uno más en el marco del cierre del seminario. En ambas ocasiones se contó con la conexión de más de cien personas y con la interacción de algunas personas participantes a través del diálogo en vivo.

02 Sobre nuestras(os) interlocutoras(es)

La selección de personas entrevistadas se hizo a sugerencia y elección grupal, una vez que se definieron los criterios de selección. Los perfiles fueron diversos, se priorizó invitar a personas que representaran o fueran activistas de diversos grupos en riesgo o en condición de vulnerabilidad, con distintas ocupaciones y que realizaran diversas actividades. En ese sentido, se buscó un equilibrio entre los lugares de enunciación desde la academia, especialistas en inteligencia artificial y los activismos.

Respecto a las personas que participaron en los talleres, éstas fueron contactadas por convocatoria abierta en las redes de las instituciones participantes, así como a través de invitaciones a colectivos y comunidades de la región sede. Previo a la impartición del taller en cada sede, las personas que asistieron respondieron el formato de encuesta. En la **Tabla 1** se muestra el perfil general de las personas que asistieron a los talleres:

Tabla 1. Perfil general de personas encuestadas/asistentes a talleres

Identidad de género	
Mujer	30
Hombre	14
No binario	4
Mujer trans	2
Hombre trans	1

Orientación sexual	
Heterosexual	23
Bisexual	9
Homosexual	8
Pansexual	5
Lesbiana	4
Asexual	2

Nivel de escolaridad	
Profesional completa	20
Posgrado	17
Profesional incompleta	6
Bachillerato	2
Secundaria	1
Primaria	1
No respondió	4

Fuente: Elaboración propia con datos recabados en la encuesta.

La zona geográfica de la que procedían las personas participantes se distribuyó de la siguiente forma: 40% procedentes de Yucatán, 31% de Aguascalientes, 23% de Baja California y 6% de la Ciudad de México. Por su parte, el nivel de educación en este grupo fue el siguiente: el 2% había completado la escuela primaria, el 4% había terminado el bachillerato, el 13% había recibido una formación profesional incompleta, el 43% una formación profesional completa y el 36% había obtenido títulos de postgrado. Dentro de la fracción que cuenta con formación profesional, los perfiles más representados fueron personas abogadas, periodistas, especialistas en salud mental, activistas, instructoras y defensoras de los derechos humanos.

03 Análisis de la información

Para explorar ampliamente el abanico de posibilidades, tomamos un grupo de las llamadas “categorías sospechosas” para escuchar un abanico rico y diverso de historias. Categorías sospechosas es el término que describe a las personas que suelen sufrir más abusos y violencia en sus comunidades, incluyendo: mujeres, personas LGBTQ+, personas indígenas y afromexicanas y personas en movilidad.[19] Este concepto, se encuentra muy relacionado con una concepción de igualdad no sólo formal ante la ley, entendida como una abstención de realizar clasificaciones prohibidas, sino con una idea de igualdad material[20] que exige del Estado (o de las organizaciones) medidas positivas. Además, esta idea de igualdad pretende mejorar la situación de estos grupos histórica o continuamente desfavorecidos.

Desde este punto de vista, esta noción de igualdad implica que el Estado no sólo evite las prácticas que potencian la marginación de estas personas, sino que también revise las normas y prácticas aparentemente neutras pero que tienen un impacto discriminatorio sobre las personas en situación de exclusión y adopte medidas positivas para ayudar a su integración en la sociedad y al acceso a los servicios sociales (por ello, se relaciona estrechamente con las llamadas “acciones afirmativas”). Por otro lado, estas distinciones históricas producen un efecto de cohesión. González Le Saux nos recuerda a Owen Fiss y su concepción de que los grupos excluidos son “grupos sociales” que “tienen su propia identidad” y “viven independientemente de sus miembros”.[21] En este mismo sentido, el grupo se caracteriza por su interdependencia, lo que implica que el escenario de sus miembros refleja el del grupo en su conjunto.

Bajo estas categorías, la información recabada tanto en las entrevistas como en los talleres, se analizó en intersección con las siguientes tres temáticas: 1) vulneración de derechos humanos: libertad de expresión, acceso a Internet, acceso a la información, privacidad y seguridad personal, equidad en redes sociodigitales, 2) agencia, y 3) generalidades y recomendaciones hacia una mayor igualdad en la garantía de la libertad de expresión y de acceso a la información.

Respecto a la encuesta, por una parte, se elaboraron estadísticos descriptivos sobre el perfil de las personas que participaron en los talleres y algunas de sus experiencias reportadas con relación a la moderación de contenido, mismas que se triangularon con la información reportada a través de las entrevistas y las experiencias compartidas durante los talleres. Estos últimos, como se mencionó previamente, formaron parte

de las actividades de intervención, pero también de investigación, al brindarnos información para enriquecer la perspectiva desde las experiencias de las personas participantes y, para proponer elementos metodológicos de intervención en esta temática.

04 Consideraciones éticas

Con el fin de resguardar la seguridad de quienes compartieron sus experiencias para esta investigación, así como de las personas asistentes a los talleres, realizamos las siguientes actividades: a) mantener el anonimato mediante la asignación de códigos de identificación; b) informarles en todo momento, de forma clara y honesta, sobre los propósitos del proyecto, sus alcances y el uso que se daría a la información recabada; y c) contar con previo consentimiento informado para cada actividad.

Un elemento a destacar es que nuestras posiciones dentro de este proyecto alternaron simultáneamente entre ser personas investigadoras, gestoras y participantes; lo anterior, lejos de operar como un sesgo, procuramos tenerlo presente al reconocer que nuestros propios conocimientos y experiencias forman parte del desarrollo de entendimientos y soluciones colaborativas. En otras palabras, consideramos que se construye una pedagogía desde la alteridad y lo colectivo, que parte del reconocimiento de nuestras subjetividades inmersas en realidades locales y al mismo tiempo globales, y en experiencias individuales que se entretejen en lo político desde un proyecto de investigación aplicada

La pretensión última, es ofrecer, desde una óptica solidaria reflexiones para la justicia y la igualdad en la sociedad digital desde una perspectiva que reconozca y sume a las voces que viven desigualdades y exclusión social. Las ideas que aquí aparecen se forjaron en comunidad y son para la comunidad. Al ser un estudio exploratorio, asumimos que de este proyecto derivan varias áreas de oportunidad y retos para futuros ejercicios de este tipo.[22]

19 Treacy, Guillermo F., “Categorías sospechosas y control de constitucionalidad”, Lecciones y Ensayos, nro. 89, 2011, Buenos Aires ps. 181-216 https://revistas-colaboracion.juridicas.unam.mx/index.php/lecciones-ensayos/article/view/13858/12368

20 González Le Saux, Marianne, y Óscar Parra Vera. «Concepciones y cláusulas de igualdad en la jurisprudencia de la Corte Interamericana. A propósito del Caso Apitz». Revista Instituto Interamericano de Derechos Humanos 1, n.o 47 (1 de enero de 2008). https://revistas-colaboracion.juridicas.unam.mx/index.php/rev-instituto-interamericano-dh/article/view/8319.

21 Fiss, Owen M. «Groups and the Equal Protection Clause». Philosophy & Public Affairs 5, n.o 2 (1976): 107-77. https://www.jstor.org/stable/2264871

22 Mohanty, Chandra Talpade. Feminism without borders: decolonizing theory, practicing solidarity. Durham:Duke University Press. 2003.

Principales resultados

La moderación automatizada puede ayudar a promover el pluralismo y la diversidad al facilitar a las personas usuarias la búsqueda de información sobre temas específicos. Sin embargo, la moderación automatizada también puede dar lugar a la violencia digital, la censura y la autocensura.

Dentro de este escenario por demás complejo, en el eje de análisis documental se identificaron las tendencias y prácticas relacionadas con la moderación automatizada de contenidos y su impacto para algunos grupos en riesgo o condición de vulnerabilidad de la sociedad mexicana, así como las posibles respuestas tanto de los individuos como de las comunidades.
El objetivo final es proporcionar enfoques individuales insertos en un análisis de lo público para entender cómo la moderación de contenidos afecta a la libertad de expresión de forma diferenciada, perpetuando con ello las distintas brechas de desigualdad.

Se han podido documentar algunos conocidos efectos positivos, como el uso de hashtags u otro tipo de alertas que ayudan a las personas usuarias a encontrar información sobre temas específicos sin tener que buscar manualmente. Sin embargo, también tiene consecuencias negativas cuando el contenido se utiliza fuera de contexto o puede ser objeto de manipulación por parte de actores políticos. Así, el estudio reiteró que la moderación automatizada puede tener impactos positivos y negativos en la libertad de expresión, dependiendo del contexto individual y colectivo; en ese sentido, no hablamos de efectos inmutables ni exclusivos.

Uno de los resultados de la presente investigación es que cada grupo estudiado tiene necesidades diferentes y se sirve de distintas redes sociodigitales para sus propios fines. Algunas comunidades indígenas, por ejemplo, tienen un uso limitado de las plataformas digitales debido a las brechas digitales existentes y la falta de conectividad a Internet en zonas rurales. Por su parte, las mujeres y personas de la comunidad LGBTQ+ sufren de intensa violencia digital y acoso. Por otro lado, las personas con discapacidad son en ocasiones estigmatizadas en las redes sociodigitales. El estudio documentó que la moderación automatizada también puede dar lugar a una forma de violencia digital, concretamente contra las mujeres y las personas LGBTQ+, que son más propensas a sufrir acoso en línea y cuyas voces tienen más probabilidades de ser censuradas.

Además, el estudio ha constatado que la moderación automatizada puede tener un efecto amedrentador sobre la libertad de expresión, ya que la gente puede autocensurarse por miedo a ser prohibida o a que se eliminen sus contenidos. Sin embargo, el presente estudio nos da cuenta de que existen (adicionalmente) factores externos que afecta relativamente este derecho como lo son: el modelo de negocio y la arquitectura de las plataformas digitales, la brecha digital y la dinámica social.

Por último, el estudio ha revelado que la moderación automatizada puede tener un impacto dispar en diferentes grupos, dependiendo de su estatus social y económico, etnicidad, género, edad y otros factores. Por ejemplo, las personas pertenecientes a pueblos indígenas son más propensas a ser censuradas debido a la barrera del idioma, mientras que personas migrantes enfrentan problemas de conectividad debido a las condiciones de tránsito migratorio, y al igual que las personas y activistas de derechos humanos son más propensas a ser objeto de acoso en línea.

En síntesis, esta investigación ofrece una visión del impacto de la moderación automatizada de contenidos en la libertad de expresión en México. Aunque los resultados son específicos para los grupos que participaron, también pueden ser relevantes para otros espacios y comunidades en riesgo o en condiciones de vulnerabilidad social.

01 Moderación de Contenidos en Plataformas Digitales

Las plataformas digitales son un ámbito de estudio muy complejo en su conceptualización por la cantidad de elementos que lo integran, sin embargo, estamos en contacto con ellas todos los días y de manera prolongada. En nuestro país, 89 millones de personas cuentan con una conexión a la Internet, de ellos su interacción en redes sociodigitales supera el 90%.[23] Prácticamente, todas las personas usuarias conviven a través de las redes sociodigitales. Los medios sociales en México se han convertido en un espacio para la difusión de información y contenidos, así como en un lugar para la interacción social y la construcción de comunidad. En este escenario, la moderación de contenidos es también el resultado de la voluntad de las plataformas digitales de proteger a las personas usuarias de los contenidos que, bajo sus normas, consideran como perjudiciales.

Este informe analiza cómo la moderación de contenidos, especialmente la automatizada, afecta a diferentes grupos de personas usuarias de medios sociales en México. Comenzaremos explicando algunos aspectos relevantes de la Internet, los sujetos que intervienen y la moderación de contenidos que se produce en las plataformas. Esto nos ayudará a entender el complejo problema que estamos tratando de abordar.

Ideas clave para el entendimiento de la moderación de contenidos

La forma más sencilla de explicar este problema es plantear las preguntas: qué es, quién lo hace y dónde, cómo y cuándo se produce la moderación de contenidos. Internet que hoy usamos tiene características muy particulares y difiere absolutamente de los inicios de esta tecnología. Sin entrar en la dimensión histórica, la Internet que hoy vivimos conecta a miles de millones de personas usuarias. De los 8 mil millones de seres humanos en el planeta, 5 mil millones han tenido conexión a Internet en los últimos

23 Asociación de Internet MX, 2022. Óp. Cit.

3 meses y se les considera usuarias; en comparación con el año 1995, en donde existían tan solo 16 millones de personas interconectadas.[24]

Las capacidades de almacenamiento, procesamiento y comunicación de información han crecido exponencialmente. De una conexión vía telefónica, tenemos ya redes intercontinentales y satelitales de capacidades extraordinarias. Esto significa que hemos pasado de computadoras gigantes y privativamente caras a super computadoras de bolsillo, que se encuentran permanentemente conectadas. El almacenamiento se ha mejorado exponencialmente, de discos magnéticos de capacidad limitada, a pequeñas tarjetas. Al mismo tiempo, la hiperconectividad ha generado ya una creciente preferencia del uso de los medios remotos en la llamada *nube*[25] frente a la compra de dispositivos para el procesamiento y almacenamiento de datos.

En consecuencia, hoy el procesamiento y el almacenamiento de nuestra información, no se lleva a cabo en nuestros dispositivos, sino en los servidores repartidos por el mundo. Cada vez más nuestras computadoras y teléfonos móviles son una pantalla conectada a Internet o bien los televisores, autos, cámaras de videovigilancia, refrigeradores, entre otros, ya están conectados a la red. Internet también se ha convertido en la suma de todas nuestras comunicaciones e interacciones sociales, una caja de pandora que refleja a la sociedad a cada interacción.

¿Qué es la moderación de contenidos?

El mismo hecho de explicar cómo suceden los fenómenos dentro de la red es un gran reto. Nuestro estudio intenta dar cuenta de una complejidad mucho mayor, incorporando la desigualdad social en un contexto tan difícil de entender como lo es esta sociedad de inicio de la segunda década del siglo XXI. Con el afán de plantear una definición mínima y una respuesta parcial a qué es la moderación de contenidos, la entendemos como “los mecanismos de gobernanza que estructura la participación en una comunidad para facilitar la cooperación y prevenir el abuso”.[26] Artículo 19 ha considerado que *la moderación de contenidos incluye los diferentes conjuntos de medidas y herramientas que usan las plataformas de redes sociales para hacer*

24 Unión Internacional de Telecomunicaciones, Estadísticas, 1996.

25 En la jerga informática, nube hace referencia a “(...) los servidores a los que se accede a través de Internet, y al software y bases de datos que se ejecutan en esos servidores [...] La nube permite a los usuarios acceder a los mismos archivos y aplicaciones casi desde cualquier dispositivo, ya que los procesos informáticos y de almacenamiento tienen lugar en servidores en un centro de datos, y no de forma local en el dispositivo del usuario”. Cloudfare. (Sin fecha). ¿Qué es la nube?. Disponible en: https://www.cloudflare.com/es-es/learning/cloud/what-is-the-cloud/

26 Grimmelmann, J., “The Virtues of Moderation”, Yale Journal of Law and Technology, EEUU, p.42.

frente al contenido ilegal y hacer cumplir sus normas de comunidad contra el contenido generado por los usuarios en su servicio[27].

Las plataformas digitales llevan a cabo el uso de estos mecanismos con el fin de controlar el contenido que se publica considerado, bajo sus términos de servicio, como nocivo o ilegal que surge diariamente en estos espacios. A este proceso de decisiones realizadas por quien controla el espacio donde se publica el contenido es lo que se le conoce como *moderación*[28].

De forma general, podemos decir que la moderación de contenido puede llevarse a cabo de dos formas: humana, por personas moderadoras, y automatizada, por mecanismos técnicos de programación o, más recientemente, herramientas de inteligencia artificial.

En la actualidad, la moderación automatizada se basa en el uso de algoritmos e inteligencia artificial para identificar, eliminar o marcar contenidos que violan las normas comunitarias de una plataforma digital. Es importante señalar que cada plataforma o proveedor del servicio cuenta con diferente naturaleza y normas comunitarias, y los procesos que utilizan para llevar a cabo la labor de moderación de forma general puede ser humana o automatizada, incluso mixta. Así, la moderación puede entenderse como "el monitoreo, evaluación, categorización, remoción [u] ocultamiento de contenido conforme a políticas de comunicación y de publicidad para propiciar las comunicaciones y comportamientos positivos y para minimizar la agresión y el comportamiento antisocial"[29]. Se trata de una definición muy amplia, pero permite tener en cuenta las diferentes estrategias de moderación de contenidos que pueden aplicarse en diversos espacios digitales como veremos más adelante.

Por otra parte, también resulta de interés el concepto de curación de contenido, que es *"el uso por parte de las plataformas de redes sociales de sistemas automatizados para clasificar, promocionar o degradar el contenido en los flujos de noticias, generalmente en función de los perfiles de sus usuarios. También se puede promocionar contenido en las plataformas a cambio de un pago. Las plataformas también pueden curar contenido mediante el uso de avisos para advertir a los usuarios sobre contenido sensible o aplicando ciertas etiquetas para resaltar, por ejemplo, si el contenido proviene de una fuente de confianza"*[30]. El contenido que vemos y no vemos depende de los algoritmos de curación de

27 Article 19. Policy: Watching the watchmen Content moderation, governance, and freedom of expression. 2021. Traducción propia. https://www.article19.org/wp-content/uploads/2021/12/Watching-the-watchmen_FINAL_8-Dec.pdf

28 Giovanni de Gregorio, en Luca Belli, Nicolo Zingales y Yasmin Curzi Terms, "Glossary Law and Policy", 2021.

29 Ibídem.

30 Artículo 19. Policy: Watching the watchmen Content moderation, governance, and freedom of expression. 2021. Óp. Cit. Traducción propia.

contenidos. Este concepto es debatido y no existe acuerdo sobre si es o no un tipo de moderación.

¿Quién realiza la moderación de contenidos?

Son muchos los actores que intervienen en los problemas que aborda esta investigación; las personas usuarias son la mayoría, pero otros actores intermediarios en las interacciones desempeñan una función crucial. La Internet requiere un gran número de instancias para funcionar con eficacia: operadores de telecomunicaciones, proveedores de nombres de dominio, empresas de alojamiento web y proveedores de servicios de Internet, son sólo algunos ejemplos de estos actores. Además, hay varias plataformas y aplicaciones disponibles que proporcionan una mezcla de servicios a los usuarios, incluidas las redes sociales, los motores de búsqueda y los proveedores de correo electrónico, entre otras.

¿En dónde ocurre la moderación de contenidos?

La Internet es un fenómeno que ocurre en el conjunto de infraestructura, es decir, el conjunto de hardware y software necesario para el funcionamiento de la red; además, constituye una red global que articula, conecta y comunica miles de redes.[31] Es importante mencionar que Internet debe estar guiado bajo el principio de gobernanza. Es decir, “es el desarrollo y la aplicación, por parte de los gobiernos, el sector privado y la sociedad civil, en sus respectivos roles, de los principios, normas, reglas, procedimientos de toma de decisiones y programas compartidos que dan forma a la evolución y el uso de Internet”.[32]

La red la podemos entender como un modelo de capas. Imaginemos una cebolla o una representación de la Tierra en la que podamos ver las capas de la atmósfera, la corteza terrestre y las capas subsecuentes hasta llegar al núcleo. En cada capa suceden fenómenos diferentes de la red y es preciso conocerla para poder explicarnos cómo interactuamos las personas. Todo ello sucede en las seis capas que conforman la Internet y que interactúan entre sí de forma jerárquica (con relación al modelo ISO y de Solum).

31 Galloway, A. (2004). Protocol: How control exists after decentralization. The MIT press. Disponible en: https://doi.org/10.7551/mitpress/5658.001.0001

32 Gobernanza de Internet, ISOC, https://www.internetsociety.org/es/learning/internet-governance/

Fuente: Elaboración propia

Contenido

Aplicaciones

Control de tráfico de datos

Internet

Enlace

Cables

Generalmente, las personas usuarias interactúan en la *capa externa* o de contenido, que corresponde a la corteza terrestre o la piel exterior de la cebolla. La siguiente capa se compone por una red informática o las *aplicaciones o programas* que se utilizan, seguida por la capa del control o transporte del tráfico de datos[33]. Las siguientes capas son las de *Internet y enlace*; y finalmente, la *capa física*, integrada por el conjunto de cables y dispositivos que hacen posible las conexiones.

En el esquema, la capa de contenido coincide con la corteza terrestre que es donde se lleva a cabo la interacción; es en esta capa en la que ocurre la moderación de contenidos. En esta capa ocurren los fenómenos de gobierno, sector privado y la interacción social. Dejaremos las fronteras marcadas para identificar, al menos, los ámbitos nacionales. Para efectos del esquema, abusaremos de su plasticidad y usaremos la atmósfera para seguir ubicando elementos en sus diferentes subcapas como los gobiernos, la sociedad civil y otros actores de lo más local a lo supranacional.

¿Cuándo ocurre la moderación de contenidos?

La moderación del contenido siempre se produce con motivo de coordinar de forma pacífica la interacción entre las personas usuarias. Puede ocurrir cuando los agentes intermediarios están estableciendo las reglas de uso de ciertos recursos, diseñando la interfaz, o cuando reaccionan a lo que hacen otras personas usuarias. Es decir, se modera contenido al establecer reglas con catálogos de contenidos no permitidos, esquemas de respuesta y sanciones, mecanismos de aplicación administrativos, técnicos y humanos, y la aplicación de estas herramientas a un determinado contenido. A este gran catálogo de conductas permitidas y no permitidas se le conoce como normas comunitarias o políticas de moderación.

En síntesis, la moderación de contenidos en las plataformas es una característica del ecosistema de gobernanza de Internet que depende de quién, dónde y cómo se apliquen estos procesos. Diferenciar el impacto de estos procesos es esencial para comprender cómo se ven afectados los diferentes actores.

¿Cómo se lleva a cabo la moderación de contenidos?

Como se mencionó líneas arriba, generalmente la moderación de contenido se lleva a cabo de dos formas: humana y automatizada. Como ejemplo podemos mencionar el caso de Facebook, en donde se remueve contenido de forma automatizada *ex-ante* como spam o pornografía infantil previo a que los usuarios puedan verlo. Por otro lado, millones de reportes llegan semanalmente sobre “probable” contenido que viola las normas comunitarias de la plataforma. En estos casos, se aplican procesos automatizados que priorizan los reportes, mismos que son analizados *ex-post* por moderadores y emiten su decisión “usualmente” en menos de 24 horas.

02 Moderación de contenidos y libertad de expresión

El análisis multidimensional es esencial al momento de examinar la moderación de contenidos, ya que permite una mirada más matizada de los efectos que esta herramienta presenta para los distintos actores. La confluencia de métodos y técnicas también nos permite identificar ámbitos en los que desarrollan los casos de moderaciones y los retos en materia de derechos que representan.

En la actualidad, el ejercicio de muchos derechos se encuentra mediado por pantallas, cables y software, por lo que es importante comprender cómo funcionan estas tecnologías para así emplearlas libre y plenamente. Cuando pensamos en la moderación de contenidos en Internet desde una perspectiva multidisciplinar, es importante considerar cómo esta práctica afecta no sólo a las personas usuarias que comparten los contenidos, sino también a quienes diseñan e implementan los mecanismos que los filtran y moderan.

La moderación de contenidos es un fenómeno complejo y polifacético que afecta a todas las partes, aunque de distinta manera. Además, ha sido ampliamente criticada por la forma en que afecta a la libertad de expresión y a otros derechos fundamentales. Desde nuestro punto de vista, la aproximación debe *aceptar la inclusión del mayor número de posiciones sociales posibles, entendidas en términos radicalmente interseccionales y bajo el supuesto de que no existe una subjetividad capaz de retomar los intereses de todas pero tampoco existe una subjetividad carente de sesgos y, por ende, la mejor forma de manejar esta variedad es al incluir la mayor cantidad de posturas dentro de un diálogo crítico radicalmente intersubjetivo*[34]. Es preciso mantener un acuerdo en el que no solo los especialistas de diversas profesiones son relevantes, sino también de las personas que utilizan, conviven y experimentan estas nuevas tecnologías. Al adoptar un enfoque de apertura, es posible desarrollar una comprensión más completa de cómo la moderación de contenidos se desarrolla y colide con los derechos de las personas.

33 Véase B. Solum, Lawrence, “Models of Internet governance”, Bygrave, Lee A. y Bing Jon (eds.), Internet Governance. Infrastructure and Institutions, New York, Oxford University Press, 2009, p. 65 y 66.

34 Guerrero Mc Manus, «Let boys be boys and girls be girls”. Una lectura crítica del concepto de “Ideología de Género” desde la Epistemología Feminista».

La dimensión de la arquitectura y diseño de Internet, nos dará los requisitos para que las reglas, mecanismos organizativos y aplicativos tecnológicos puedan convivir en el ecosistema. De estos saberes, podemos rescatar elementos valiosísimos para diseñar procesos de moderación coordinados y descentralizados que tengan en cuenta la pluralidad de contextos culturales, lingüísticos y sociales en los que se aplicarán[35].

La forma en la que se toman las decisiones en Internet no debe pasar desapercibida. La gobernanza de Internet, en específico la autoridad para tomar decisiones sobre el diseño, funcionamiento y uso de Internet es policéntrica[36], lo que ha generado un cierto grado de desconcierto en la forma de percibir las relaciones de poder. El modelo de participación de múltiples actores es uno de los aspectos más relevantes de la gobernanza de Internet, ya que permite un proceso de toma de decisiones más horizontal y descentralizado. Ejemplo de estos son los foros de toma de decisiones de la ICANN (Internet Corporation for Assigned Names and Numbers), que se caracterizan por ser un espacio de participación en el que se toman decisiones sobre la gestión técnica y normativa de Internet.

Ahora bien, cuando nos planteamos cómo hacer que las plataformas de redes sociales logren potenciar y proteger la libertad de expresión y otros derechos, comprendemos que que el modelo de autorregulación de las plataformas digitales o las regulaciones restrictivas por parte de los estados en relación a la moderación de contenido no es suficiente para garantizar tales derechos[37]. Las acciones de estas plataformas tienen un profundo impacto en nuestras sociedades, y es necesario establecer mecanismos que den certeza en el ejercicio de los derechos humanos asociados al uso de redes socio digitales.

¿Por qué es importante la libertad de expresión en la Internet?

Los derechos humanos son de todas las personas y garantizan su dignidad, independientemente de quiénes sean, dónde residan, cuál sea su identidad, a qué se dediquen, etcétera. Se encuentran generalmente contenidos en las constituciones de los países, en sus leyes y en el Derecho Internacional de los Derechos Humanos. Actualmente, las posibilidades de conexión a Internet posibilitan el

35 Feeney, M., 2021. Twitter Is Just the Beginning of Jack Dorsey's Speech Revolution, Cato Institute. Retrieved from https://policycommons.net/artifacts/2019788/twitter-is-just-the-beginning-of-jack-dorseys-speech-revolution/2772240/ on 25 Jun 2022. CID: 20.500.12592/x9vtj6.

36 Scholte, J. A. (2017). Polycentrism and democracy in internet governance. The net and the nation state: Multidisciplinary perspectives on Internet governance, 165-184.

37 Watching the watchmen, Content moderation, governance, and freedom of expression, Article 19, 2021, pp. 17-18. https://www.article19.org/wp-content/uploads/2021/12/Watching-the-watchmen_FINAL_8-Dec.pdf

ejercicio de distintos derechos, incluidos los derechos humanos, independientemente de las fronteras geográficas. Además, las plataformas digitales se han convertido en un espacio para la difusión de información y fomento de la participación ciudadana.

No obstante, Internet también ha sido empleado para violar los derechos de ciertas personas o grupos; al respecto, grandes operadores han sido criticados por su papel en la difusión de discursos de odio e información falsa. En función de lo anterior, diversos actores gubernamentales y no gubernamentales como son los organismos internacionales y las organizaciones de la sociedad civil se han pronunciado por la necesidad de garantizar los derechos humanos en Internet. Entre ellos se encuentran, por ejemplo, la Organización de las Naciones Unidas[38], la Comisión Interamericana de Derechos Humanos[39] y varios gobiernos nacionales.

Reconociendo que las empresas proveedoras de espacios digitales tienen la responsabilidad de garantizar que sus plataformas no se utilicen para afectar el libre ejercicio de los derechos humanos, en mayo de 2016 la Comisión Europea en acuerdo con Facebook, Microsoft, Twitter y YouTube firmaron el *Code of conduct on countering illegal hate speech online* con el fin de ayudar a las personas usuarias a notificar discursos de odio ilegales en las plataformas digitales, así como mejorar la coordinación entre éstas y las autoridades nacionales[40]. De igual forma, las empresas han tomado algunas medidas para intentar solucionar estos problemas, como aumentar la transparencia en torno a los anuncios políticos como en el caso de México con el Instituto Nacional Electoral[41].

Asimismo, existen pronunciamientos de las Naciones Unidas y de autoridades del Sistema Interamericano de Derechos Humanos en las que se incluye a las empresas como sujetos directamente responsables en la promoción y protección de los derechos humanos, no sólo en su propio territorio, sino también en el extranjero.

Estos derechos se encuentran también reflejados en los términos de uso y en las normas comunitarias. Sin embargo, es importante recordar que la moderación de contenido por parte de empresas privadas no se produce en un vacío legal; por el contrario, estas empresas deben garantizar que sus prácticas de moderación de contenido se ajusten a las normas internacionales de derechos humanos[42]. Si bien, en la moderación se privilegia a veces la privacidad u otro derecho, la consecuencia limita la capacidad de

38 Véase Relator Especial de las Naciones Unidas (ONU) sobre la Promoción y Protección del derecho a la Libertad de Opinión y de Expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), también en UNESCO, Global toolkit for judicial actors. International legal standards on freedom of expression, access to information and safety of journalists. 2021; así como Naciones Unidas (ONU). Representante Especial del Secretario General para la cuestión de los derechos humanos y las empresas transnacionales y otras empresas. Principios Rectores sobre las empresas y los derechos humanos: puesta en práctica del marco de las Naciones Unidas para 'proteger, respetar y remediar" (A/HRC/17/31)

39 CIDH, Relatoria Especial para la Libertad de Expresión, Libertad de expresión e Internet, OEA/Ser.L/V/II CIDH/RELE/INF.11/13, 31 diciembre 2013, párr. 87; , Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión, y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). 1 de junio de 2011, Principio 3. C. Disponible en: https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=849&lID=2

40 https://ec.europa.eu/newsroom/just/items/54300

41 Central Electoral. «Facebook e INE anuncian colaboración para elecciones», 5 de febrero de 2018. https://centralelectoral.ine.mx/2018/02/05/facebook-e-ine-anuncian-colaboracion-para-elecciones/ .

42 "La autorregulación, cuando es efectiva, sigue siendo la forma más adecuada de abordar los problemas profesionales relacionados con los medios de comunicación. De acuerdo con el principio 9 de los Principios de Camden, todos los medios de comunicación deben, como responsabilidad moral y social y a través de la autorregulación, desempeñar un papel en la lucha contra la discriminación y la promoción de la comprensión intercultural, incluidas las siguientes consideraciones: (a) Tener cuidado de informar en el contexto y de una manera objetiva y sensible, al tiempo que garantizar que los actos de discriminación se pongan en conocimiento del público. (b) Estar atentos al peligro de fomentar la discriminación o los estereotipos negativos de individuos y grupos en los medios de comunicación. (c) Evitar las referencias innecesarias a la raza, la religión, el género y otras características de grupo que puedan promover la intolerancia." OHCHR. «OHCHR | The Rabat Plan of Action». Accedido 25 de junio de 2022. https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action. Traducción propia.

una persona para expresarse y de todas las personas para tener acceso a esa información[43].

Tanto personas usuarias como intermediarias están obligadas a seguir las reglas establecidas en acuerdos y a través de la autorregulación, pero los derechos humanos permanecen como un marco de validez y la legitimidad de estas decisiones. Hoy día Internet es una herramienta que, se espera, pueda ser utilizada por cualquier persona también para ejercer sus derechos[44]. Uno de estos derechos es la libertad de expresión, mismo que se ha entendido como una fibra esencial del tejido democrático. En los últimos años, este derecho ha cobrado aún más relevancia en la era digital. El uso de las redes sociales y otras plataformas en línea ha dado un nuevo significado a la libertad de expresión, ya que se ha convertido en una herramienta esencial para el ejercicio de este derecho.

Como puede deducirse, el ejercicio de diversos derechos articula con lo digital y se sitúa en los marcos de la libertad de expresión. Por ejemplo, el derecho a la privacidad es esencial para el libre desarrollo de la personalidad, ya que permite a las personas elegir la información que comparten y con quién. Asimismo, la libertad de reunión y de asociación son también derechos fundamentales que se ejercen en línea. El honor, la honra y la reputación también están en juego en el mundo digital, lo cual pone en riesgo el derecho de toda persona a que se proteja su integridad y a impugnar las informaciones falsas o engañosas sobre ella. Aunado a lo anterior, no puede omitirse mencionar el incremento en los asesinatos a periodistas y personas defensoras de los derechos humanos, así como la persecución de que suelen ser objeto tanto dentro como fuera del espacio digital.

El pleno ejercicio de la libertad de expresión en México y en buena parte de América Latina está en peligro. Por todo ello, las empresas deben garantizar que los contenidos se retiren o bloqueen sólo cuando sea necesario y proporcionado para lograr un objetivo legítimo en virtud de la legislación internacional sobre derechos humanos.

43 Lanza, E. y Matías, J., Moderación de Contenidos y Mecanismos de Autorregulación. “El Oversight Board” de Facebook y sus implicancias para América Latina, El Diálogo, 2021.

44 Relatoría Especial para la Libertad de Expresión, Comisión Interamericana de Derechos Humanos, OEA, Open Society Foundations, Libertad de expresión e Internet, diciembre 2013, p. 6. https://www.oas.org/es/cidh/expresion/docs/informes/2014_04_08_Internet_WEB.pdf

¿Por qué la moderación de contenidos está relacionada con la libertad de expresión?

En primer orden, la moderación de contenidos implica siempre una decisión entre un contenido y una consecuencia sobre éste. La posibilidad de reducir su relevancia, retirarlo de la vista pública, eliminarlo o incluso suspender o eliminar la cuenta del servicio pueden ser las consecuencias. Por lo tanto, tendrá un impacto en la capacidad de expresión de la persona usuaria y una limitación en el alcance al que puede llegar el contenido. Ya sea que otra persona usuaria, una autoridad o la propia plataforma detone estas consecuencias, el hecho es que tendrán un impacto relevante para la libertad de expresión. Las obligaciones en materia de derechos humanos son sustantivas, pero también procedimentales. Es fundamental recordar que el tamaño y el número de personas usuarias de las plataformas deben corresponder a los procedimientos de moderación que utilizan, lo mismo acontece con sus responsabilidades en materia de derechos.

En un primer nivel, las reglas relacionadas con el contenido deben ser acordes a los estándares internacionales en materia de derechos humanos. Es decir, deben ser claras, proporcionadas y no discriminatorias. Además, estas normas deben ser conocidas de antemano por las personas para que puedan tomar decisiones informadas sobre el uso de la plataforma.

Por su parte, los mecanismos a través de los que se toman decisiones deben tener ciertas características en su procedimiento. En general, estas decisiones deben ser transparentes, predecibles, motivadas y sujetas a revisión; además, la persona usuaria debería tener siempre la posibilidad de presentar su versión de los hechos y, si es necesario, impugnar la decisión ante un organismo independiente.

Las empresas privadas que ofrecen plataformas de redes sociodigitales son responsables de garantizar que estas plataformas sean respetuosas con los derechos humanos, incluido el derecho a la libertad de expresión. En este sentido, deben poner en marcha sistemas y procedimientos que les permitan tomar decisiones de forma transparente, predecible, motivada y sujeta a revisión.

Además, las empresas privadas a menudo están en una posición de poder sobre los y las usuarias, incluso con alguna relevancia o labor gubernamental. Por ello, deben asegurarse de que sus normas y procedimientos de moderación de contenidos no den lugar tratos diferenciados, promoción de prejuicios u otras formas de desigualdad[45]. La situación es significativa porque las empresas privadas se comprometen a evitar ciertas conversaciones como los discursos de odio y estigmatizantes, y la pornografía infantil entre otros delitos.

El hecho de que las empresas privadas se encarguen de moderar los contenidos no significa que puedan actuar sin escrutinio[46]. Al contrario, estas empresas deben rendir cuentas de sus actos. En este sentido, es esencial que las personas usuarias tengan acceso a recursos efectivos en caso de que su derecho a la libertad de expresión sea comprometido. En nuestra perspectiva, ante la falta de debidos controles, existen potenciales disminuciones sustantivas a las capacidades de comunicación y de alcance de nuestra información relacionadas con la moderación de contenidos que representan un reto enorme para usuarios, gobiernos, empresas y mecanismos de protección de Derechos Humanos.

¿Hay censura en las redes socio digitales?

En términos generales se puede decir que propiamente no hay censura en las redes sociodigitales. Cuando el Estado censura contenidos, suele hacerlo a través de sus leyes, reglamentos y procedimientos. Por ejemplo, en algunos países sancionan jurídicamente a quienes critican al gobierno o difunden información sobre ciertos temas[47]. En estos casos, el Estado puede bloquear sitios web o cuentas de redes sociales que no cumplan con sus normas. Sin embargo, los gobiernos también llegan a utilizar las redes sociodigitales para difundir su propaganda y censurar las voces críticas[48,49]. Para ello, recurren a diferentes tácticas, como el bloqueo de personas usuarias o contenidos, la manipulación de la información o el uso de las herramientas de vigilancia para identificar a las disidencias[50]. También, mediante orden administrativa o judicial emitida por una autoridad competente, los gobiernos pueden requerir a las plataformas digitales que impongan sanciones o eliminen cierto contenido[51].

45 ONU, Informe de la Alta Comisionada de las Naciones Unidas para los Derechos Humanos acerca de los talleres de expertos sobre la prohibición de la incitación al odio nacional, racial o religioso. 2013. Disponible en: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G13/101/51/PDF/G1310151.pdf?OpenElement

46 CIDH, Relatoria Especial para la Libertad de Expresión, Libertad de expresión e Internet, OEA/Ser.L/V/II CIDH/RELE/INF.11/13, 31 diciembre 2013, párr. 87; y Relator Especial de las Naciones Unidas (ONU) sobre la Promoción y Protección del derecho a la Libertad de Opinión y de Expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión, y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). 1 de junio de 2011, Principio 3. C. Disponible en: https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=849&lID=2

47 Committee to Protect Journalists. (Sin fecha). Los 10 países con la mayor censura. Disponible en: https://cpj.org/es/2015/04/los-10-paises-con-la-mayor-censura/

48 Artículo 19. (2020). #LibertadNoDisponible. Censura y remoción de contenido en Internet Caso: México. Disponible en: https://articulo19.org/wp-content/uploads/2021/02/LIBERTAD-NO-DISPONIBLE-single-page.pdf

49 UN Human Rights. (8 de marzo de 2017). UN experts urge States and companies to address online gender-based abuse but warn against censorship. Disponible en: https://www.ohchr.org/en/press-releases/2017/03/un-experts-urge-states-and-companies-address-online-gender-based-abuse-warn

50 R3D. Gobierno Espía. Vigilancia sistemática a periodistas y defensores de derechos humanos en México. 2017. Disponible en: https://r3d.mx/wp-content/uploads/GOBIERNO-ESPIA-2017.pdf

51 Cantoral, Karla. Daño moral en redes sociales: su tratamiento procesal en el derecho comparado. Rev. IUS [online]. 2020, vol.14, n.46 pp.163-182. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-21472020000200163&lng=es&nrm=iso>

La censura no es solo un mecanismo estatal para silenciar cierta información, en la práctica los agentes privados también pueden participar en ella[52]. Las redes sociodigitales pertenecen a empresas privadas que pueden decidir sobre qué contenidos se permiten en sus plataformas; esta actividad "entre privados" supone un riesgo al derecho a la libertad de expresión y al acceso a la información. Sin embargo, esto no implica que las empresas estén por encima de la ley: dependiendo de su origen, las empresas tienen obligaciones jurídicas y sociales, en las que generalmente se encuentran sujetas a la obligación de respetar una serie de derechos; por lo que son consideradas responsables de sus actos y omisiones.

La censura en las redes socio digitales puede adoptar diferentes formas: desde la creación de las normas comunitarias restrictivas, el establecimiento de instancias humanas para la revisión de contenidos que carecen del contexto de la comunicación, hasta el uso de algoritmos sesgados que filtran y eliminan automáticamente los contenidos legítimos[53,54,55]. Así, las empresas privadas pueden decidir eliminar los contenidos que consideran inapropiados o contrarios a sus condiciones de uso. Si las decisiones se encuentran fuera de un entendimiento mínimo de los estándares de derechos humanos, podemos afirmar cierta forma de censura. En algunos casos, las empresas privadas pueden incluso cerrar las cuentas de las personas usuarias.

Además de los agentes privados y el Estado, las personas usuarias -de forma individual o colectiva- pueden generar actos de censura, por ejemplo, denunciando contenidos que consideren inapropiados. En algunos casos, esto puede llevar a la eliminación de contenidos o incluso al cierre de cuentas de usuario en actitudes como el *digital mobbing*[56].

Como se mencionó líneas arriba, la censura tiene efectos directos en la persona que la sufre y trascienden a las distintas áreas de su vida. Además, puede tener un efecto amedrentador sobre quienes presencian o conocen la censura, ya que puede disuadirles de compartir determinados contenidos por miedo a las represalias que han observado en otras(os). Ante estas consecuencias negativas, muchas personas usuarias deciden autocensurarse, esto significa que deciden no compartir ciertas informaciones u opiniones por miedo a sufrir consecuencias negativas. Como se analizará en el apartado 5 de este documento, para personas y grupos en situación

52 Comisión Interamericana de Derechos Humanos. [CIDH]. (2000). Declaración de principios sobre libertad de expresión.

53 Consejo asesor de contenido. (Junio de 2022). Oversight Board overturns Meta's original decision in 'Reclaiming Arabic words' case (2022-003-IG-UA). Disponible en: https://oversightboard.com/news/428883115451736-oversight-board-overturns-meta-s-original-decision-in-reclaiming-arabic-words-case-2022-003-ig-ua/

54 Greenspan, R. y Tenbarge, K. (26 de septiembre de 2020). YouTuber's channels and videos are being mistakenly deleted for debunking COVID-19 conspiracy theories. Insider. Disponible en: https://www.insider.com/youtube-demonetized-covid-19-disinformation-moderation-automation-bots-strikes-2020-9

55 Llansó, E. J. (2020). No amount of "AI" in content moderation will solve filtering's prior-restraint problem. Big Data & Society, 7(1), 2053951720920686.

56 Actividad consistente en utilizar Internet para herir o atemorizar a otra persona, especialmente mediante la transmisión de comunicaciones desagradables: Las palabras inglesas "digital mobbing", que incluyen el acoso en Internet, el ciberacoso y el ciberacoso, se utilizan para referirse a una variedad de formas de difamación, intimidación y coerción llevadas a cabo utilizando tecnologías electrónicas como salas de chat, programas de mensajería instantánea. Oben Uru, F. The Dark Side of Digitalization: Digital Mobbing. In F. Ozsungur (Ed.), Handbook of research on digital violence and discrimination studies. Business Science Reference. 2022.

de riesgo o vulnerabilidad -como es el caso de las personas LGBTQ+, periodistas, migrantes, activistas o ciertos grupos de mujeres- la autocensura puede ser una forma de autocuidado, no obstante, se trata de una decisión individual resultado de prácticas externas de coacción, amenaza y acoso, entre otras formas de violencia.

¿Cuáles son los efectos de la moderación automatizada de contenido?

La libertad de expresión debería ser una característica integrada en cualquier diseño por defecto, sin embargo, el empleo de algoritmos para la evaluación de contenidos podría tener un impacto significativo en este derecho debido a que las soluciones de moderación se basan con frecuencia en tecnologías de inteligencia artificial que no se encuentran libres de criterios sesgados y discriminatorios. Como resultado, pueden afectar de forma desproporcionada a determinados grupos de personas usuarias, como algunas mujeres, personas indígenas y afromexicanas, personas con alguna discapacidad o personas pertenecientes a la comunidad LGBTQ+.

Así mismo, la aplicación de la moderación automatizada de contenido puede dar lugar a la eliminación de información fidedigna. Esto se debe a que estas tecnologías no siempre son capaces de identificar correctamente el escenario en el que se produce la comunicación[57]. Si bien existen algunas reglas sobre contenidos claramente no permitidos, deberían existir en el mismo sentido, etiquetas al contenido especialmente protegido.

¿Cómo se relacionan la libertad de expresión y la no discriminación?

La libertad de expresión es un derecho humano que incluye la libertad de buscar, recibir y difundir información e ideas de todo tipo, sin importar las fronteras. Por su parte, el principio de no discriminación establece que todas las personas son iguales, pero reconoce que dicha igualdad se viola de forma agravada cuando, por la existencia de prejuicios sobre una persona o grupo, se produce su exclusión, marginación e impedimento para el libre ejercicio de derechos[58].

57 Haimson, O., Delmonaco, D., Nie, P., y Wegner, A. (2021). Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-35.

58 Courtis, citado por González, M. y Parra, O. (2008). Concepciones y cláusulas de igualdad en la jurisprudencia de la Corte Interamericana. A propósito del Caso Apitz. Revista IIDH. P. 135.

Ambos conceptos están estrechamente relacionados, ya que la libertad de expresión sólo puede garantizarse si no se discrimina a las personas por prejuicios basados en su raza, género, clase social, edad, orientación sexual, o cualquier otra condición. En otras palabras, si se excluye a determinados grupos de personas del debate público, no se está respetando la libertad de expresión[59]. En específico, la censura se contrapone a la libertad de expresión y al principio de no discriminación, ya que restringe la información a la que la gente puede acceder y limita las voces que pueden ser escuchadas en el debate público aunado a la protección especial sobre el discurso relativo a las identidades en el marco jurídico interamericano[60].

En el contexto de las redes sociodigitales, la libertad de expresión debe garantizarse de forma que no se discrimine a ningún grupo o persona usuaria. Esto significa que las políticas de las plataformas y las prácticas de moderación de contenidos no deben tener un impacto desproporcionado en ningún grupo.

¿Existen modelos de regulación de moderación de contenidos?

No existen modelos de regulación de moderación de contenidos sino sistemas de protección de derechos humanos nacionales e internacionales. En áreas como la protección de datos personales, la diferencia en el entendimiento de la privacidad como un derecho constitucional o no, ha generado modelos como el americano, europeo y los mixtos[61]. En el ámbito de la moderación de contenidos en las redes sociodigitales, es posible idear diferentes modelos según el modelo de protección y límites de la libertad de expresión de cada región.

La búsqueda de un modelo ideal sería aquel en el que las plataformas tendrían la obligación de retirar los contenidos que violen los derechos de las personas usuarias, pero también podrían permitir la divulgación de cierto tipos de contenidos que podrían ser ofensivos si es que existe un interés público en mantenerlos accesibles. El reto es ingente para todas y todos los actores involucrados, pues implica repensar y rearticular su relación con el fenómeno y buscar soluciones colectivas.

59 Salazar Ugarte, Pedro y Gutiérrez Rivas, Rodrigo, El derecho a la libertad de expresión frente al derecho a la no discriminación. Tensiones, relaciones e implicaciones, México, UNAM, Instituto de Investigaciones Jurídicas-Consejo para Prevenir la Discriminación, 2008.

60 Comisión Interamericana de Derechos Humanos. Marco jurídico interamericano sobre el derecho a la libertad de expresión. New York; CIDH; Asdi: OEA ;, 2010.

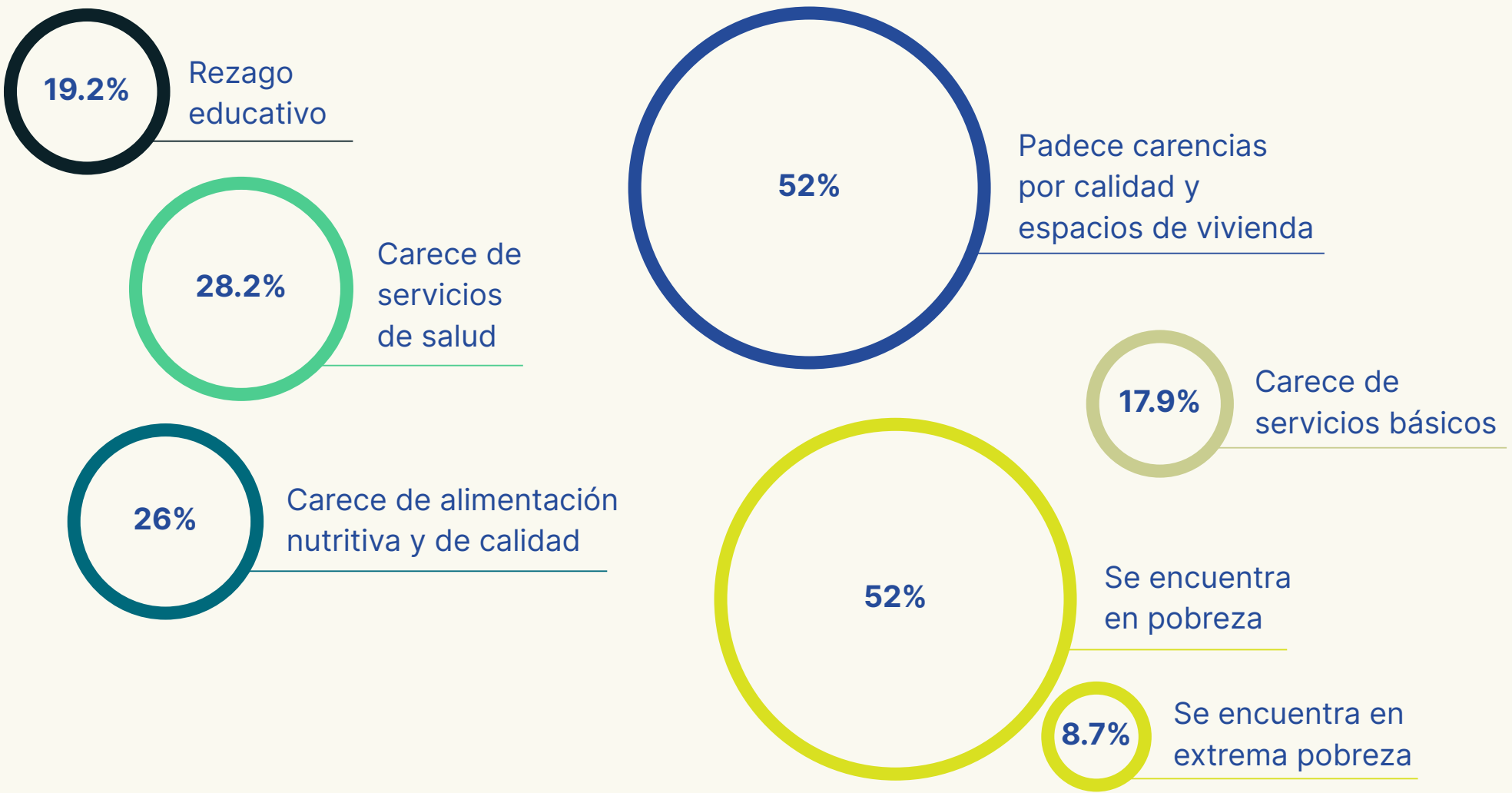
61 Montezuma, L. The case for a hybrid model on data protection/privacy. International Association of Privacy Professionals, May 6, 2020. ttps://iapp.org/news/a/the-case-for-a-hybrid-model-on-data-protectionprivacy .

03 México en y fuera de línea. Indicadores y contexto

Algunos indicadores básicos

En México viven más de 126 millones de personas -51% son mujeres y 49% son hombres-, cuyas condiciones de vida son variadas y, en muchos casos, se encuentran atravesadas por condiciones de desigualdad (**Imagen 2**). Así mismo, la diversidad social y cultural es notable: por mencionar algunos datos, 6% de la población habla una lengua indígena, 2% se identifica como persona afromexicana, 77.7% se adscribe a la religión católica, mientras que el 8.1% se declara sin adscripción religiosa.[62]

Imagen 2. Cifras sobre desigualdad en México (2020)



Aunado a lo anterior, las personas inmigrantes en México suman 1 212 252, la mayoría proveniente de los EEUU[64].Nuestro país es un importante país de tránsito y de destino de personas migrantes, muchas de ellas desplazadas forzadas con necesidad de protección internacional como es el caso de personas provenientes de la mayor concentración, de Guatemala, Honduras, El Salvador en segundo lugar y de Venezuela entre otros. México es además un importante país de origen y retorno de migraciones desde Estados Unidos; es decir, es una nación *transterritorial* con más de once millones de personas nacidas en México residiendo en el país del norte, por lo que se calcula que una población total de 160 millones de personas mexicanas al incluir a quienes habitan fuera del país[65].

Personas, comunidades y tecnologías

En cuanto a la presencia de las tecnologías de la información en los hogares, en el año 2020, el 44% de la población declaró tener una computadora. Se trata de un aumento considerable en comparación con el 35% del año 2013, pero es inferior al pico del 46% de 2016. Las computadoras son sin lugar a dudas las herramientas que diversifican en mayor medida el uso de la Internet por las posibilidades que representan[66].

Mientras que en 2013 tan solo el 30% de los hogares tenía conexión a Internet en México, para 2020 esta cifra se elevó al 60%. Para el mismo año, el 72% de la población ya utilizaba Internet, lo cual supone casi un 30% más que en 2013[67]. Dado el contexto de confinamiento social derivado la pandemia por COVID-19, 2020 fue un año clave en el incremento del uso de Internet desde los hogares, reportando que solo el 6% de las personas internautas no utilizó Internet en casa (cifra que para 2013 fue de 44%).

En lo que respecta a las redes sociodigitales, casi el 30% de la población en México de 15 años o más utiliza estos medios para mantenerse informada de lo que ocurre en su comunidad. Esta cifra aumenta hasta el 43% en el caso de las personas de 20 a 29 años, y cae hasta el 10% para el grupo de 60 años o más. Por su parte, el 70 por ciento de la población que dice estar interesada en los asuntos públicos se informa a través de la televisión, el 44,7% la obtiene de los medios sociales (redes sociodigitales en términos de este documento) y el 22% de otras fuentes en Internet[68].

Fuente: Elaboración propia con datos de CONEVAL (2020)[63]

62 Instituto Nacional de Geografía y Estadística [INEGI]. (2020). Censo de Población y Vivienda 2020. Datos disponibles en: https://www.inegi.org.mx/programas/ccpv/2020/#Documentacion

63 Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), Resultados de pobreza en México 2020 a nivel nacional y por entidades federativas, 2020. Disponible en: https://www.coneval.org.mx/Medicion/Paginas/PobrezaInicio.aspx

64 INEGI, Censo de Población y Vivienda 2020, Óp. Cit.

65 Guillén, T. México, nación transterritorial. El desafío del siglo XXI , Universidad Nacional Autónoma de México, Programa Universitario de Estudios del Desarrollo, 2021.

66 Instituto Nacional de Geografía y Estadística, Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2020. Disponible en: https://www.inegi.org.mx/programas/dutih/2020/

67 Ibídem.

Recientemente, la Asociación de Internet MX presentó datos correspondientes a 2021, entre los cuales destaca que en México existen 89.5 millones de internautas, lo cual equivale al 75.7% de la población[69]. La distribución de esta población se encuentra bastante equilibrada en cuanto a sexo con un 51% de usuarias mujeres y 49% de usuarios varones. En lo que respecta a la distribución por nivel socio económico, el 41.6% son de los grupos D y E, 35.8% a los grupos C y C-, mientras que el 13% corresponde al nivel C +. El grupo más aventajado corresponde al 8.8% conjuntando a los niveles A y B[70].

Estos marcadores permiten analizar uno de los impactos diferenciados en lo que refiere a la brecha digital: el acceso. Desafortunadamente, una cuarta parte de las personas en México están excluidas de Internet[71]. Además, para el caso de este grupo poblacional, la brecha digital no se refiere únicamente al acceso sino al tipo de uso que se da a las tecnologías de la comunicación. No obstante, quienes están desconectadas corresponden en su mayoría a los grupos socioeconómicos más desfavorecidos, concentrando casi el 90% de los casos. Además del sexo, la residencia urbana o rural determina otra gran brecha, siendo las áreas rurales las que más déficit de conectividad presentan[72]. Asimismo, la edad es un factor relevante, pues casi la mitad de las personas desconectadas son mayores de 55 años[73]. En síntesis, la brecha digital por desconexión está concentrada en la población más pobre, menos educada y de mayor edad.

El tipo de conexión es importante y su distribución entre los segmentos analizados nos da elementos interesantes para nuestro análisis. La distribución de personas usuarias entre redes fijas, celulares y su combinación, nos permite ver que los grupos A y B se encuentran permanentemente conectados en un 97%. Este dato se reduce casi al 70% en el grupo C +, pero cae al 51% en los grupos D y E. Existe una brecha sobre la disponibilidad de conexión, puesto que mientras algunas personas se encuentran permanentemente en línea, otras tienen conexiones de acuerdo con el lugar en el que se encuentren. Las líneas móviles de banda ancha siguen siendo privativamente caras.

Las plataformas de redes sociodigitales son utilizadas primordialmente desde el teléfono móvil con un 96.7%. Sobre los servicios y marcas utilizadas, para videollamadas, el 63% usa

Fuente: Elaboración propia con datos de la encuesta.

68 Instituto Nacional de Estadística y Geografía (INEGI), Encuesta Nacional de Cultura Cívica (ENCUCI) 2020, disponible en: https://www.inegi.org.mx/programas/encuci/2020/

69 Asociación de Internet MX. 18° Estudio sobre Hábitos de los Usuarios de Internet, Óp. Cit.

70 La Asociación Mexicana de agencias de Inteligencia de Mercado y Opinión (AMAI) clasifica los hogares en 7 niveles socioeconómicos (NSE) con base en seis características del hogar que son: escolaridad del jefe(a) del hogar, número de dormitorios, número de baños completos, número de personas ocupadas de 14 años o más, número de autos y contar con internet fijo en la vivienda. Los siete NSE, de mejor calidad de vida a menor son: A/B, C+, C, C-, D+, D y E. Para más detalle sobre los NSE se sugiere consultar el siguiente sitio web: https://www.amai.org/NSE/index.php?queVeo=preguntas

71 Asociación de Internet MX, 2022. Óp. Cit.

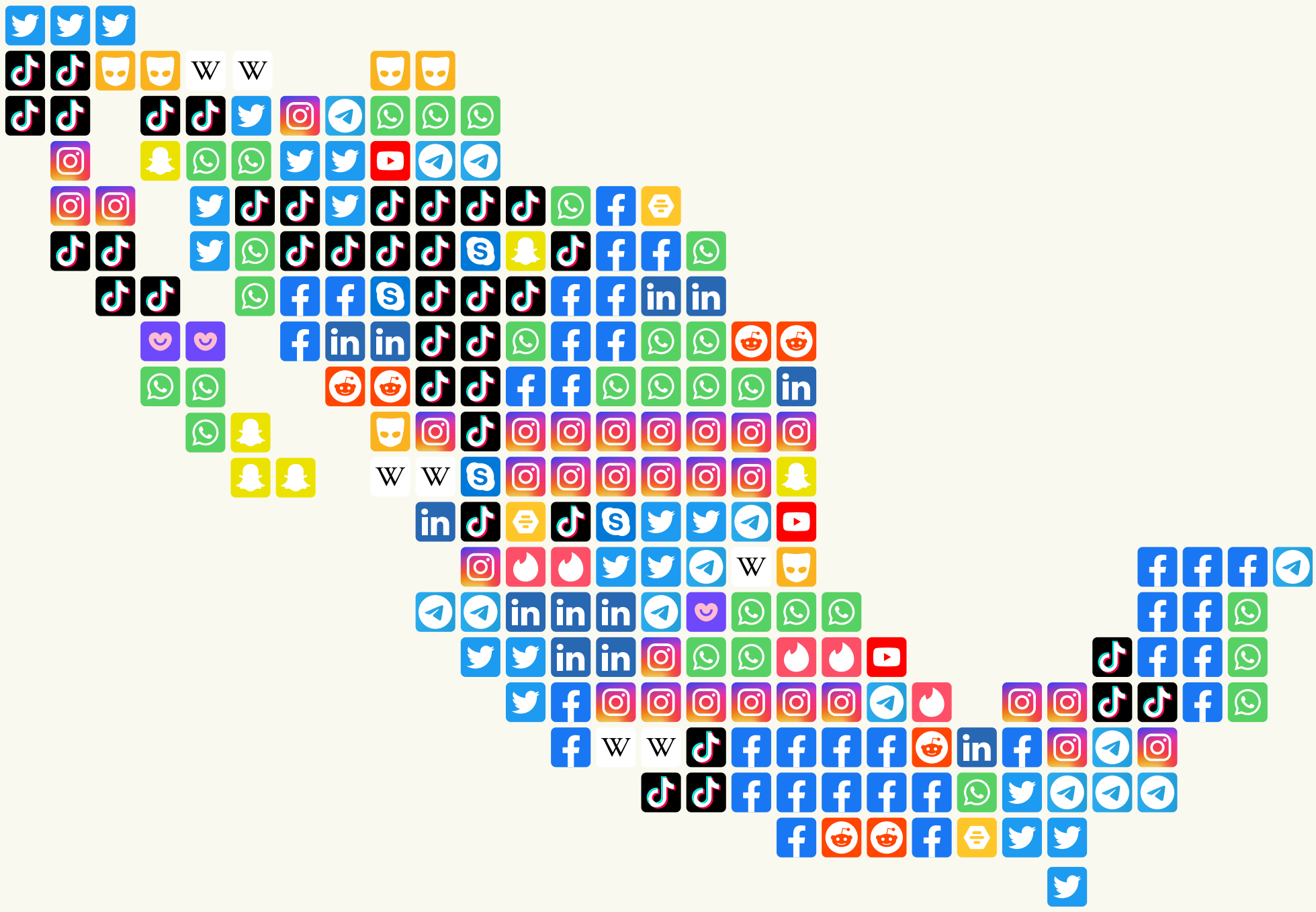
72 Article 19. (2022). Negación. Informe anual 2021. Disponible en: https://articulo19.org/wp-content/uploads/2022/04/Book-1_A19_2021_V03_BAJA-3.pdf

73 Asociación de Internet MX. 18° Estudio sobre Hábitos de los Usuarios de Internet, Óp. Cit.

74 Asociación de Internet MX, 2022. op. cit.

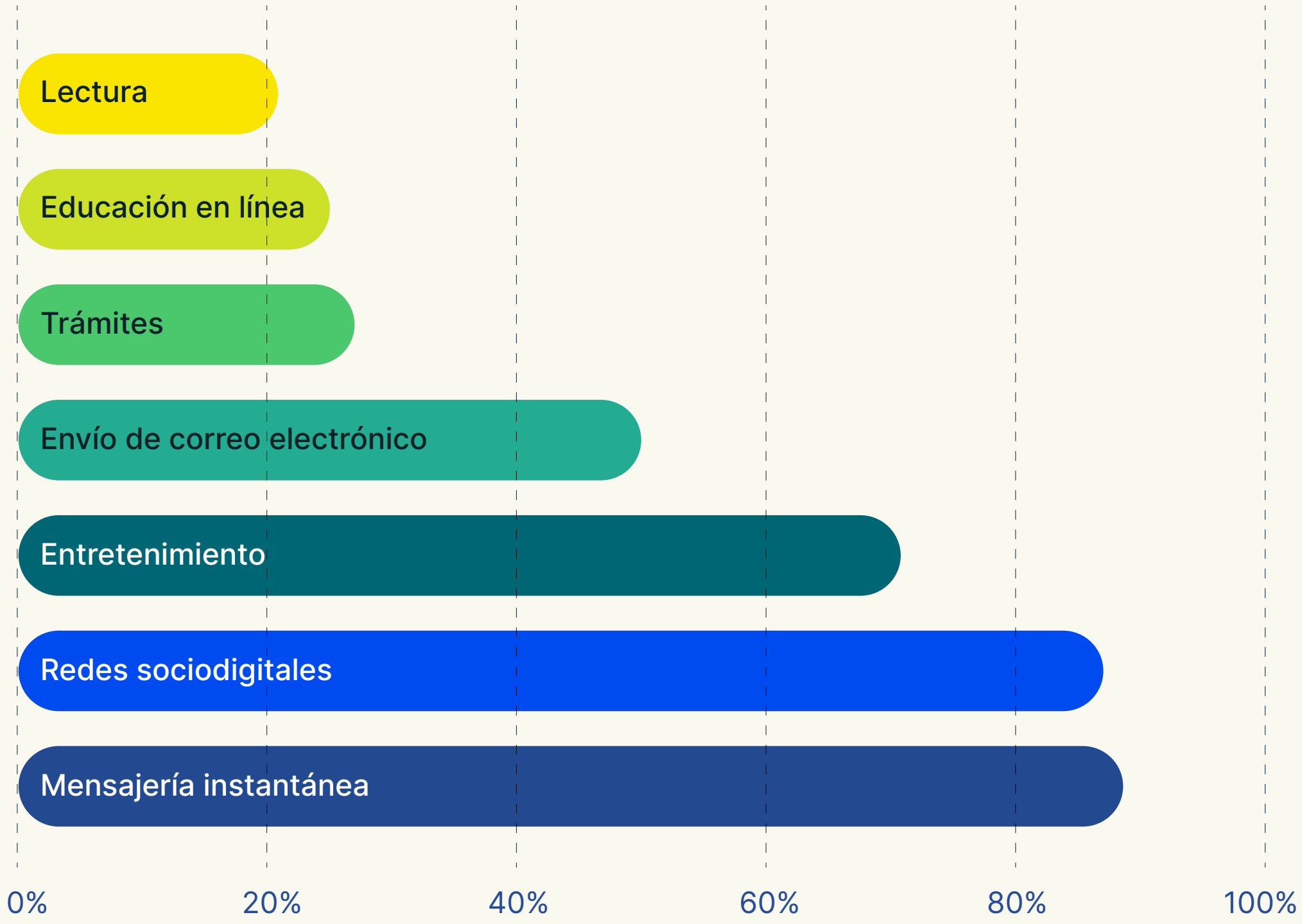
WhatsApp, el 46% Zoom, 27% Google Meet, 18% Facebook y el 9.7% Microsoft Teams. Meta se lleva la concentración de conocimiento de las personas usuarias mexicanas, ya que WhatsApp, Facebook e Instagram rebasan sustancialmente a sus competidores en conocimiento. Se acercan YouTube de Alphabet, Twitter y TikTok, siendo esta última la de mayor crecimiento en fechas recientes subiendo al top 5 en solo 2 años[74]. Nuestra herramienta de encuesta arrojó datos similares, siendo casi las mismas redes con mayor número de menciones:

Imagen 3. Redes sociodigitales más empleadas



Respecto a las actividades que se realizan cuando se está en línea, las actividades que tienen que ver directamente con el ejercicio de ciertos derechos como la educación no representan los porcentajes más altos (**Gráfico 1**), sin embargo, la comunicación y permanencia en redes sociodigitales se vinculan indirectamente con el derecho a la información[75].

Gráfico 1. Actividades realizadas mientras se está en línea



De las personas encuestadas para el presente estudio, un 94% manifestó usar redes sociodigitales diariamente, el 4% declaró usarlas “cuando tienen tiempo” y el 2% no usa redes. Sobre el tiempo de conexión, el 37% declaró usarlas menos de 5 horas, el 37% de 5 a 10 horas y el 27% siempre están conectadas[76].

En tanto a los riesgos reportados, las personas usuarias de plataformas digitales refirieron el robo de datos (68%), recepción de virus en el dispositivo (37%), invasión a la privacidad (29%), recepción de contenido inadecuado (24.7%), recepción de noticias falsas (19%) y ciberacoso (9.2%)[77]. La encuesta aplicada para este estudio planteó una serie más abierta de riesgos en línea, encontrando que el 19% de las personas encuestadas ha sufrido un bloqueo por contenido inapropiado, al 9% le han bloqueado su cuenta por incumplimiento de las normas comunitarias, y el 44% ha sufrido la eliminación de contenidos por la misma razón. Aunque mínimos, encontramos cierres de cuentas en un 5% para los casos individuales, y 3.5% para las cuentas de organizaciones[78].

Respecto a prácticas de violencia, del total de personas encuestadas para este estudio, el 36% afirmó haber sido objeto de discursos estigmatizantes o de odio, y un 42% ha sufrido mensajes privados ofensivos o amenazas. Otras situaciones reportadas fueron la extorsión con un 14% y la suplantación de identidad con 13%. De quienes recibieron alguna sanción por parte de alguna plataforma, 21% declaró no haber obtenido respuesta por parte de esta sobre el por qué de la sanción.

Como es en línea, es fuera de ella. México, sentimientos y resentimientos

En seguimiento al análisis de las personas mexicanas y su relación con estas tecnologías, nos ocupa identificar los rasgos autoritarios y discriminatorios que potencialmente pueden verse reflejados en línea. Existen diversos ejercicios estadísticos para la medición de estos elementos, tales como los realizados por el Instituto Nacional de Geografía y Estadística (INEGI), la Universidad Nacional Autónoma de México, el Instituto Nacional Electoral, y el Consejo Nacional para Prevenir la Discriminación (CONAPRED), entre otras instituciones. De estos estudios, el más reciente es la Encuesta Nacional de Cultura Cívica (ENCUCI) 2020 y uno de los más detallados es la Encuesta Nacional sobre Discriminación 2017 (ENADIS), ambos documentos referencia para este informe[79].

Fuente: Elaboración propia con datos de Asociación de Internet MX.

75 Asociación de Internet MX. 18° Estudio sobre Hábitos de los Usuarios de Internet, Óp. Cit.

76 Datos de la encuesta levantada para este estudio.

77 Asociación de Internet MX, 2022. Óp. Cit.

78 Idem.

79 Se han revisado la Encuesta Nacional de Identidad y Valores, la Encuesta Nacional de Cultura Política, la Tercera Encuesta Nacional de Cultura Constitucional, la Encuesta Nacional de la Sociedad de la Información, así como la Encuesta Nacional de Ciencia y Tecnología.

De los documentos citados, los valores y métricas varían para la descripción de la percepción de algunos elementos sustantivos de la vida en sociedad. Uno de los elementos persistentes es la identificación de valores y actitudes cívicas. Son también relevantes las percepciones sobre prejuicios, estigmas y estereotipos.

Para recortar nuestro objeto de estudio, los valores que estimamos relevantes son la confianza y la tolerancia en varias dimensiones. Por otro lado, resultan importantes las percepciones sobre casos de discriminación, así como las redes de contención con las que cuentan las personas.

04 Una perspectiva desde las experiencias: voces diversas y puntos de encuentro

Al tratarse de tecnologías desarrolladas y empleadas por personas, Internet y las redes sociodigitales no son espacios neutrales. Por supuesto, es innegable el impacto positivo que han tenido para visibilizar las necesidades y derechos de grupos históricamente marginados, al tiempo que se han convertido en espacios de enunciación importantes para estas poblaciones. Sin embargo, estas bondades no se han manifestado de forma homogénea para todos los grupos ni para sus integrantes, así como tampoco se han dado libres de obstáculos. En ese sentido, aun cuando exista el común denominador de que son poblaciones cuyo derecho a la libertad de expresión se encuentra en constante amenaza, la forma en que ésta se experimenta adquiere matices.

Así, por ejemplo, de acuerdo con lo reportado por las personas entrevistadas para esta investigación, sabemos que la brecha en el acceso a Internet resulta más amplia para la población indígena, lo cual, aun tratándose de una desventaja en términos de contar con un espacio de enunciación, también implica que no suele ser atacada directamente en espacios digitales. No obstante, también refleja y confirma cómo la limitante en el acceso produce obstáculos para acceder a información, fortaleciendo así cadenas de desigualdad y acumulación de desventajas. En ese sentido, cabe señalar que ninguna de las categorías de opresión opera de forma aislada. Por tanto, aun cuando el punto de vista de las personas entrevistadas resalta su experiencia desde una condición, sus realidades deben pensarse a la luz de la complejidad de las distintas posiciones que ocupan en el entramado social. Por ejemplo, las exclusiones que enfrentan las mujeres indígenas en el acceso y uso de plataformas digitales no son las mismas que enfrentan los varones jóvenes indígenas.

Aunado a lo anterior, es importante señalar que los impactos tanto positivos como negativos que estos grupos han experimentado

en el espacio digital no se encuentran desvinculados de lo que les acontece fuera de él. De hecho, como bien señalan las y los estudiosos de lo digital[80], hoy día la vida social acontece entre conexiones y tránsitos[81], y el espacio digital constituye una extensión del espacio público[82]. Por tanto, es indispensable dimensionar que los resultados que a continuación se presentan no se limitan a lo que las personas experimentan al formar parte de, por ejemplo, una red sociodigital, ni tampoco los efectos de las exclusiones y violencias de que son objeto se limitan a dicho espacio.

A continuación, presentamos la información aportada por las personas participantes en la encuesta, las entrevistas y en los talleres, misma que hemos organizado en torno a cinco puntos: a) moderación automatizada de contenido, b) libertad de expresión, c) inequidades en el acceso a Internet y a la información, y d) seguridad personal, salud mental y agencia. Finalmente, se retoman las aportaciones de nuestras(os) interlocutoras(es) para alimentar las recomendaciones del apartado final de este documento.

Moderación automatizada de contenido

Un elemento que destaca en la narrativa tanto de las personas entrevistadas como de quienes respondieron la encuesta y participaron en los talleres es que abordar algún contenido político o social en redes sociodigitales implica enfrentarse directa o indirectamente a algún tipo de restricción. En ese sentido, la regulación de contenido se torna discriminatoria en sí misma, ya que se refleja en el silenciamiento de unas voces frente a la promoción de otras. Si bien, menos del 20% de las personas encuestadas reportaron que en al menos una ocasión alguna plataforma les bloqueó un contenido por considerarlo inapropiado o por infringir normas comunitarias, esto podría tener relación con que no han subido todo el contenido que desearan, es decir, la autocensura es consecuencia de la moderación de contenido y genera la ilusión de que las plataformas no remueven tanto contenido, ya que las personas usuarias lo hacen por anticipado; es decir, para evitar que la plataforma baje un contenido, la persona usuaria no lo sube. No obstante, lo anterior puede entenderse también como formas estratégicas para permanecer en las redes, tal y como se mencionó en el taller de Aguascalientes (realizado el 22 de febrero de 2022)

Fuente: Elaboración propia con datos de la encuesta

80 Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., y Tacchi, J. (2019). Etnografía digital. Ediciones Morata.

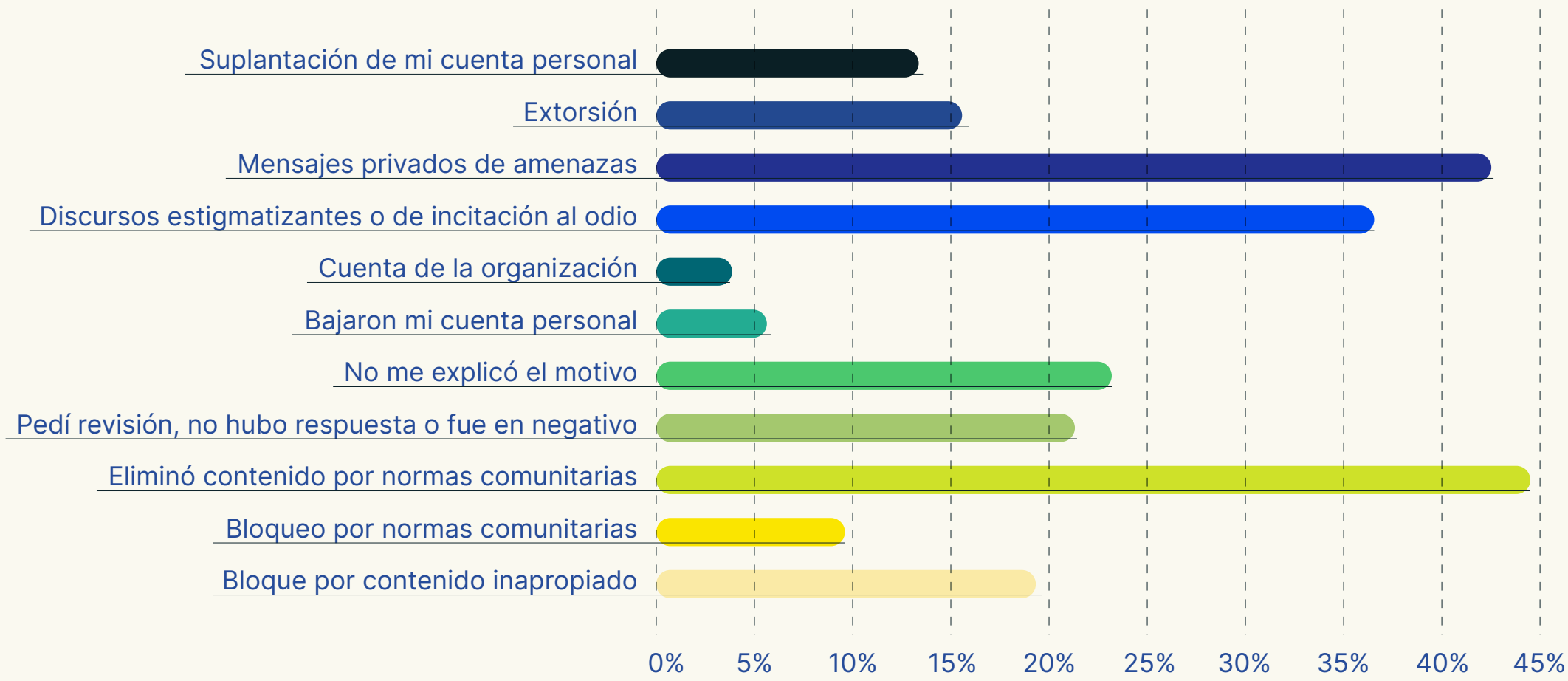
81 Gutiérrez, A. (2016). Etnografía móvil: una posibilidad metodológica para el análisis de las identidades de género en Facebook. Revista interdisciplinaria de estudios de género de El Colegio de México, 2(4), 26-45.

82 Calderón, F., y Castells, M. (2019). La nueva América Latina. Fondo de Cultura Económica.

respecto a no subir fotos personales cuando se considera que se es víctima potencial de violencia o usar cuentas alternas para subirlas.

Como parte de las situaciones reportadas por quienes fueron personas interlocutoras vinculadas con moderación de contenido, destaca la eliminación de contenido bajo la justificación de haber infringido normas comunitarias (**Gráfico 2**). Seguidos de ésta, se observa la exposición a discursos que incitan al odio y la recepción de mensajes amenazantes. Estas dos situaciones se relacionan con la moderación, ya que, pese a tratarse de contenido que vulnera la dignidad y seguridad de las personas, no fueron removidos.

Gráfico 2. Experiencias relacionadas con moderación de contenido



A partir de su experiencia con remoción de contenido, varias de las personas entrevistadas concluyen que los algoritmos dedicados a la moderación son selectivos, no solo porque no aplican criterios neutrales reales, sino porque carecen de criterios para evaluar caso por caso. Esta selectividad participa en que ciertos grupos mantengan su privilegio de enunciación, mientras que otros no puedan sostenerlos. Al respecto, las personas entrevistadas señalan prácticas relacionadas con dicha selectividad:

* Limitar las acciones de una persona usuaria dentro de alguna red sociodigital (acción conocida como *baneo*). Estas limitaciones son más frecuentes para personas defensoras de derechos humanos; en cambio, es menos frecuente para quienes difunden discursos de odio u otro tipo de contenido como, por ejemplo, pornografía infantil.

* Favorecer el bloqueo de cuentas de personas de usuarias que utilizan palabras altisonantes (ejemplo: “pinche hetero”), no así de quienes comparten frases discriminatorias, transfóbicas o sexistas, por nombrar algunas (ejemplo: “hombres con falda”). Así, aun cuando las redes sociodigitales no son las que crean los contenidos, sí toman decisiones con consecuencias, sobre lo que las personas producen.

* Se da prioridad por salvaguardar las cuentas que pagan por publicidad.

En este mismo sentido, las personas entrevistadas coinciden en la importancia de controlar lo que se difunde en redes, sin embargo, la ausencia de criterios eficientes para ello se torna altamente problemático. Ejemplos de lo anterior es la dificultad para bloquear un comentario homofóbico que, al mismo tiempo, cuenta con millones de “me gusta”, o el de retirar los discursos patologizantes y xenófobos que se producen alrededor de la población migrante. Desde la parte técnica, el control es complicado por la enorme cantidad de personas usuarias.

Una apreciación compartida por las personas entrevistadas es que las redes sociodigitales no cuentan con la capacidad para gestionar de forma equilibrada los debates, razón por la que es frecuente observar que terminan en polarización. De acuerdo con la persona entrevistada 07, especialista en Internet y moderación de contenido, el fenómeno de la polarización se presenta con menor fuerza en redes que dividen sus discusiones por temas (tal sería el caso de Reddit) debido al factor de *autoselección*, es decir, las personas interactúan deliberadamente con otras a quienes son relativamente afines. Aunado a ello, este tipo de redes cuentan con moderadores humanos, a diferencia de redes como Facebook, donde la moderación queda a cargo de algoritmos. En ese sentido, los problemas asociados a la polarización descansaría en la autoselección y en la *escala*, es decir, la cantidad de personas que

interactúan y la capacidad del agente moderador para gestionar dicha interacción.

Desde otra perspectiva, las personas especialistas en tecnología consultadas para este estudio, consideran que la polarización de la sociedad ha existido antes e independientemente de Internet, de hecho, muchas veces se gesta desde fuera; el problema radica en que las redes sociodigitales la amplifican por lo que en términos publicitarios resulta redituable. Así mismo, la polarización también ha provocado que temáticas relativas a la democracia se conviertan en un producto de entretenimiento. Otro problema es que no admite matices en el debate, desconociendo –y hasta sancionando– los diversos contextos en que se presenta un fenómeno y las distintas formas para comprenderlo.

Con base en lo anterior, se considera que la supuesta neutralidad en la moderación automatizada de contenido corresponde más bien a una ficción. La ausencia de criterios éticos en la moderación automatizada impacta desde la permisividad de *discursos antiderechos*[83] hasta en la posibilidad de verificar una cuenta. Al respecto, la persona entrevistada 06, integrante de una asociación civil que combate el racismo en México, refiere que no han podido verificar la cuenta de Twitter de la organización, dado que el nombre incluye la palabra “racismo” y, suponen, el algoritmo de la red considera que se trataría de una cuenta que fomenta el racismo. Por tanto, los criterios para evaluar si una palabra es discriminatoria o se trata de un insulto no se aplican contextualmente, y en la práctica, implica que se normalice la circulación de discursos de violencia hacia ciertos grupos. Otro punto relacionado con el tema del lenguaje es que, para el caso de los *hashtags*, éstos logran mayor difusión si contienen palabras que formen parte del banco de datos de la red en que se difunden.

Otra situación identificada por varias de las personas entrevistadas es que los algoritmos de las redes sociodigitales facilitan que las publicaciones de personas defensoras de derechos humanos lleguen a grupos antiderechos. Lo anterior, fomenta que sus publicaciones sean denunciadas y sus cuentas canceladas temporal o permanentemente. Perder la oportunidad de enunciarse en el espacio digital desequilibra la difusión de discursos, además de, por supuesto, vulnerar el derecho a la libre expresión.

83 Coincidimos con Cruz, “Los llamamos grupos antiderechos debido a que su lucha pretende suprimir derechos específicos, como a la interrupción legal del embarazo, al matrimonio de personas del mismo género, al libre desarrollo de la personalidad e identidad, a la adopción por personas del mismo género y el derecho de las mujeres a llevar una vida distinta a la que marca el estereotipo de género, acciones que realizan en defensa de un orden que presentan como natural y que se condensan en la promoción de un modelo de familia enmarcado por una constelación de valores determinados” Cruz, D. «Afecto, cuerpo e identidad: reflexiones encarnadas en la investigación feminista». En Rabasa y Guerrero Mc Manus, Afecto, cuerpo e identidad. Reflexiones encarnadas en la investigación feminista, 238., editado por Alba Pons Rabasa y Siobhan Guerrero Mc Manus. Serie Estudios jurídicos, núm. 330. Ciudad de México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas, 2018.

Por otra parte, la persona entrevistada 01, investigadora y académica, señala otra cara de la moderación que se relaciona con que las redes sociodigitales no explican con claridad sus lineamientos de operación, lo cual lleva a que haya personas usuarias que interpretan las restricciones como actos punitivos –cosa que no siempre es así– y no como regulaciones necesarias para la convivencia. En consecuencia, la remoción de contenido puede producir lógicas de rencor y venganza que se traducen en más violencia.

Libertad de expresión

Las personas entrevistadas coincidieron en que la libertad de expresión ejercida en las redes sociales sin respeto y controles adecuados, ha sostenido la presencia de ciertas voces y opacado la emergencia de otras. Es decir, las plataformas y los actores que intervienen en la moderación de contenidos actúan desde parámetros que consideran neutrales y objetivos, sin embargo, muchas veces derivan en la amplificación de ciertos colectivos y la supresión de otras voces. Lo primero no es problemático en sí, pero el desequilibrio de voces perpetúa la inequidad y promueve silenciamientos. La libertad de expresión vinculada con la idea de democracia fomenta una tendencia a que “todas las personas hablen”, sin embargo, surge la pregunta de si existen las condiciones para que todas las personas hablen y sean consideradas sujetos de interlocución.

Tras el análisis de los datos se observa que el derecho a la libertad de expresión suele emplearse como justificación para producir y viralizar discursos ofensivos y/o de odio[84] contra grupos históricamente discriminados y contra personas activistas. Entre las prácticas dentro de redes sociodigitales identificadas por las personas entrevistadas que directa o indirectamente han atentado contra su libertad de expresión individual y de los grupos con quienes colaboran se encuentran:

01 **Discursos de odio y discriminación.** Todas las personas asistentes a los talleres refirieron en la encuesta estar de acuerdo en que las redes sociodigitales son espacios potencialmente inseguros donde no se respetan las opiniones, menos aún cuando estas provienen de voces marginales. La

mediación de la pantalla conduce a la deshumanización y esto, a su vez, facilita violentar. El discurso de odio constituye un discurso discriminatorio y violento contra una persona o grupo por cuestiones de su identidad, género, etnicidad, raza u opinión, entre otras, y se caracteriza por incitar "[...] a lastimar la dignidad e integridad física de las personas con posibilidades reales e inminentes de que el daño se materialice"[85]. La proliferación de estos discursos resulta más visible en redes como Facebook y Twitter. Cada que una persona usuaria replica estos mensajes genera una suerte de respaldo, aun cuando, en ocasiones, dicho mensaje proviene de perfiles falsos. Por su parte, su potencial viralización hace suponer que los grupos antiderechos se encuentran organizados y en constante vigilancia de lo que se dice en redes sociodigitales. Si bien, los discursos de odio no pueden tipificarse como ilegales, las personas entrevistadas refieren que eso no es justificación para que se difundan; además, su masiva difusión participa en la normalización del odio y la violencia en general. Además de las redes sociodigitales, es en los sitios web de medios de comunicación donde pueden encontrarse comentarios racistas, clasistas, sexistas, etcétera, emitidos por las personas usuarias. En lo referente a poblaciones indígenas, dichos comentarios suelen aparecer con frecuencia en las notas periodísticas sobre megaproyectos de alto impacto ambiental.

02 **Estigmatización.** En conexión con el punto anterior, los discursos de odio contribuyen a perpetuar los estigmas que pesan sobre los grupos históricamente discriminados; a su vez, dichos estigmas se emplean como justificación para violentarles y restringirles la posibilidad de enunciación tanto dentro como fuera de redes.

03 **Intención de homogeneidad.** Un fenómeno que atenta contra la libertad de expresión es la intención –consciente o inconsciente– de imponer una perspectiva (opinión, posición, actitud, etcétera) como la verdadera y correcta, por lo que aquellas que no coinciden se interpretan como erróneas y susceptibles de ser desdeñadas. La homogeneidad se promueve en redes sociodigitales dado que presentan contenido individualizado con el que simpatizan las personas, creando así una especie de endogamia discursiva. La falta

84 Para evaluar si se trata de un discurso de odio, se requiere la presencia y análisis de los siguientes elementos: 1) contexto socio-político-económico en que se encuadra el discurso, 2) nivel de alcance, 3) probabilidad de que el discurso se traduzca en un daño hacia una persona o grupo, 4) posición y grado de influencia de quien emite el discurso, 5) propósito deliberado de causar un daño, y 6) contenido que incita a la violencia. Article 19. (7 de agosto de 2020). ¿Qué es el discurso de odio? Disponible en: https://www.youtube.com/watch?v=j7QjhlQNbPU

85 Ibíd.

de convivencia con la diversidad provoca que cuando las personas entran en contacto con opiniones distintas a las propias, se muestren reticentes al diálogo, más aún cuando dichas opiniones atentan contra el *status quo* o provienen de grupos discriminados socialmente. La pretensión de homogeneidad impide que las redes sociodigitales sean espacios democráticos, pues considera a algunos sujetos-grupos solo como objeto de discusión y no como sujetos interpelables.

04 **Denuncia y bloqueo de páginas, perfiles y contenido.** Estas prácticas tienen motivaciones diversas en lo específico, pero en lo general responden a la pretensión de homogeneidad. En lo específico, activistas por los derechos de las personas con discapacidad refieren que se les han denunciado publicaciones por corregir el uso correcto de términos para referirse a la discapacidad. En conjunto, quienes fueron entrevistadas consideran que la corrección de información falsa desata, en general, una actitud defensiva en un buen número de personas usuarias. Las denuncias en ocasiones son masivas, es decir, las y los usuarios se unen colectivamente para denunciar una página o perfil. Así mismo, las denuncias suelen ir acompañados de acusaciones y *linchamiento* en redes; además, muchas provienen de *bots*. Cuando la denuncia deriva en el bloqueo del perfil/página o en la remoción de contenido, esto representa una baja en el flujo de visitas, reduciendo así los espacios de enunciación. En lo que respecta a los contenidos visuales, los motivos por los que son removidos o señalados negativamente varían. En el caso de la población trans, las fotografías de sus cuerpos suelen ser retiradas de algunas redes; en contraste, las imágenes de personas con alguna discapacidad, migrantes o indígenas no son removidas, pero suelen ser interpretadas como prácticas de revictimización, por lo que algunas personas se expresan para que no sean compartidas. Como puede observarse, los discursos disidentes (textuales y gráficos) son más susceptibles de ser removidos que aquellos que no lo son. Este trato diferenciado constituye una forma de discriminación que mantiene las brechas de desigualdad en el ejercicio del derecho a la libertad de expresión y produce un panorama sesgado de la realidad social.

05 **Acoso y violencia.** Estas prácticas se manifiestan de distintas formas. En ocasiones, se expresan como agresiones directas hacia personas y, en otras, hacia colectivos o asociaciones que promueven la defensa de sus derechos. En el caso de las mujeres, destacan las agresiones a aquellas que se adscriben como feministas y que están a favor de la legalización del aborto. Sobre las personas trans, la temática que suele despertar comentarios violentos es el de las infancias trans. Respecto a las personas migrantes, se producen discursos xenofóbicos. Sobre la población indígena, las agresiones se dirigen más hacia las personas defensoras del territorio a través de discursos racistas y clasistas. Finalmente, las personas que experimentan alguna discapacidad suelen ser interpeladas desde la condescendencia, la caridad y la lástima.

Como puede observarse, la experiencia de las personas entrevistadas refiere obstáculos importantes para el ejercicio de su derecho a la libertad de expresión. Dichas limitantes son compartidas, pero adquieren particularidades según la población. Destaca que, como se abordará más adelante, el acceso limitado a Internet de las poblaciones indígenas y migrantes provoca que su confrontación directa con discursos de odio sea limitada, en este caso, suelen ser más las personas activistas y defensoras de derechos quienes son objeto de esas agresiones y de la denuncia de sus perfiles. Por su parte, las poblaciones con discapacidad ven restringida su libertad de expresión debido al tabú que existe alrededor de ellas; la interlocución que se tiene con esta población suele darse desde la caridad y la condescendencia, por lo que mucho de su contenido –especialmente de tipo visual– se interpreta como revictimizante. Finalmente, las poblaciones LGBTQ+ y de algunos grupos de mujeres, tanto sus integrantes como las asociaciones y colectivos que participan en la defensa de sus derechos son blancos directos de agresiones dentro de las redes sociodigitales, presentando una alta denuncia de sus perfiles y publicaciones.

En conjunto, las amenazas a la libertad de expresión en el espacio digital dan cuenta de un marco deficiente de derechos humanos en México que perpetúa las desigualdades y se articula con la vulneración de otros derechos para estas poblaciones. En ese sentido, resulta necesario no solo perfeccionar los marcos institucionales y de las propias plataformas, sino también producir

espacios de reflexión colectiva donde la población en general genere mayor conciencia sobre la importancia de contar con normas comunitarias desde una perspectiva de derechos humanos.

Inequidades en el acceso a Internet y a la información

Además de la libertad de expresión, el derecho a la conectividad y a la información disponible en Internet también presenta limitantes para los grupos analizados. En lo que respecta a la conectividad, nuestros(as) interlocutores(as) de comunidades indígenas refirieron que suelen tener un acceso limitado lo cual implica, como se mencionó previamente, que las probabilidades de que experimenten agresiones directas en espacios digitales sean menores, sin embargo, también conlleva, similar a lo reportado por otros estudios[86,87], a que el ejercicio de su derecho al acceso a la información se vea obstaculizado.

De acuerdo con las personas entrevistadas, el limitado acceso a Internet responde, en parte, a que las empresas que proveen el servicio no consideran rentable insertarse en las zonas alejadas. Si bien, el Estado ha generado dependencia a Internet para la realización de trámites y para recibir apoyos monetarios, a la par no se ha ocupado de garantizar la conectividad. En ese sentido, la conectividad universal resulta una ficción, así como el pensar que ciertas poblaciones, como el caso de los pueblos indígenas, no gestionan apoyos del gobierno por desinterés, cuando en realidad responde más a impedimentos externos.

Para quienes sí acceden a Internet, su uso se ve limitado en ocasiones, ya que no todas las personas indígenas manejan el idioma español ni la lectura de su lengua originaria (suponiendo que accedan a contenido en lengua indígena). No obstante, el acceso limitado no significa que esta población se encuentra desvinculada de lo digital. Por una parte, hay quiénes sí utilizan redes sociodigitales, pero su uso se dirige más a fines de compra-venta. Por otro lado, las personas entrevistadas para este grupo refieren que la población indígena tejen redes digitales a través de terceros, como puede ser las asociaciones civiles y personas activistas.

El acceso diferenciado de conexión a Internet ejemplificado líneas

86 Article 19. (2022). Negación. Óp. Cit. Pp. 56-64 y 69.

87 Article. (2020). COVID. Article 19 Oficina para México y Centroamérica. Pp. 25.

arriba se conoce como *brecha digital*, la cual tiene un impacto directo en el ejercicio de ciertos derechos como el de acceso a la información, por ello es considerada como una amenaza a la igualdad, la democracia y la justicia social[88].

La brecha digital se puede definir como una distribución diferenciada en el acceso, manejo e integración a la vida cotidiana de las TICs[89] que, a su vez, refleja las diferencias sociales, económicas, culturales, raciales y étnicas al interior de la sociedad[90]. La carencia de infraestructura para conectarse a Internet o de conocimiento sobre el uso de las tecnologías y sus dispositivos, son dos de las expresiones de la brecha digital[91].

La conectividad debe evaluarse tanto numéricamente (cuántas personas cuentan con conexión), como cualitativamente, es decir, examinar si el acceso permite a las personas satisfacer sus necesidades[92]. En ese sentido, se habla de una *conectividad significativa*, la cual se compone de cuatro elementos: i) uso regular a Internet (que las personas acceden diariamente); ii) dispositivo apropiado (que las personas cuenten con los dispositivos necesarios); iii) datos suficientes (contar con datos ilimitados) y iv) velocidad adecuada de la conexión (velocidad suficiente para satisfacer la demanda de las y los usuarios)[93]. Así mismo, como señalan algunas de las personas entrevistadas, debe considerarse el uso que se da a lo digital, es decir, evaluar si las personas cuentan con el conocimiento suficiente para sacar el máximo provecho de las plataformas digitales.

Otra de las poblaciones que destaca por un acceso a Internet y a dispositivos digitales más limitado son las personas migrantes, especialmente cuando se encuentran en tránsito. Las personas entrevistadas señalaron que estas restricciones son estructurales, es decir, anteceden su arribo a México, pues las viven desde su país de origen. Dentro de esta población, son las mujeres e integrantes de la comunidad LGBT+ en quienes la brecha de acceso es más amplia. Dentro de los medios digitales más empleados por estas poblaciones está el servicio de mensajería instantáneo Whatsapp. Este medio suele ser usado para difundir información sobre trámites, acceso a servicios y rutas migratorias. Al respecto, una problemática identificada es que muchas veces la información que se comparte no es verídica, pero es altamente compartida y creída porque resulta esperanzadora. La difusión de esta información no necesariamente

88 United Nations [UN]. (1995). Conferencia Mundial sobre la Mujer. Disponible en http://cedoc.inmujeres.gob.mx/documentos_download/100073.pdf

89 Gómez, D., Alvarado, R., Martínez, M. y Díaz, C. (2018). La brecha digital: una revisión conceptual y aportaciones metodológicas para su estudio en México. Entreciencias: diálogos en la sociedad del conocimiento, 6(16), 47-62. Disponible en:https://www.redalyc.org/journal/4576/457654930005/html/

90 Berrío, C., Marín, P., Ferreira, E. y Chagas, E. (2017). Desafíos de la inclusión digital: antecedentes, problemáticas y medición de la brecha digital de género. Psicología, Conocimiento y Sociedad, 7(2), 121-151. Disponible en http://www.scielo.edu.uy/pdf/pcs/v7n2/1688-7026-pcs-7-02-00121.pdf

91 Mendoza, J. y Caldera, J (2014). Umbrales para la determinación de la brecha digital: comparativa entre regiones desarrolladas. Transinformação, 26, 125-132. Disponible en https://www.scielo.br/j/tinf/a/v6gZqZHYCYFzmnXvmdMCSCJ/?lang=es

92 Secretaría de Gobernación. (2021, abril). Acuerdo por el que se da a conocer el Programa de Conectividad en Sitios Públicos 2020-2021 de la Secretaría de Comunicaciones y Transportes. Recuperado el 10 de enero de 2022 de: http://dof.gob.mx/nota_detalle.php?codigo=5616105&fecha=16/04/2021

93 Ídem.

es mal intencionada, sino que responde a la falta de información pública y la reactividad de las acciones gubernamentales que suelen responder a las coyunturas políticas, por lo que generan incertidumbre, desconcierto y poca transparencia.

Relativo al tema de la información, se señala también que las instituciones gubernamentales no siempre producen y/o almacenan información suficiente y de calidad para estas poblaciones y, mucho menos, en idiomas que no sea español. En ese sentido, la libertad de expresión no sería el problema, sino la calidad y cantidad de la información que se comparte en dicho ejercicio de libertad. Un ejemplo de ello lo experimenta la población LGBT+ sobre la que se observa un retorno de discursos biologicistas y patologizantes. Sumado a lo anterior, cuestionar la calidad de la información suele interpretarse como intento de censura, lo que deriva en que las personas que cuestionaron la calidad terminen censuradas.

Otros problemas identificados por las personas entrevistadas respecto a la información que se difunde sobre estos grupos en situación de vulnerabilidad es, por un lado, que pocas veces se acompaña de una reflexión ética. Además, se suele responsabilizar a las personas usuarias de consumir información falsa en lugar de cuestionar la carencia de mecanismos necesarios para que las instituciones que sí cuentan con la información real, la hagan llegar a la población. En cuanto a la desinformación , cabe decir que también llega a difundirse a través de espacios de defensa de derechos humanos a modo de información sensacionalista. Esto obedece a que, en ocasiones, no hay interés en buscar información, sino en adoptar un discurso que tenga popularidad y provea autocomplacencia. Es decir, hay quienes priorizan que un mensaje se viralice frente a que sea verídico.

Además de la brecha en el acceso, otra de las dificultades que enfrenta la población indígena y migrante es la autocensura, misma que limita el flujo de información que las personas comparten y reciben. Estas comunidades tienden a no publicar fotos en redes sociodigitales por una creencia generalizada de que hacerlo es revictimizante y, en el caso de migrantes, para no correr riesgo de ser identificadas y detenidas. De igual forma, asociaciones que apoyan a la población migrante se ven limitadas en publicar fotos del interior de los centros de retención, pues les implicaría problemas legales. Por otra parte, en redes circula información

errónea sobre la población con alguna discapacidad, la cual va desde la lástima hasta el fomento del porno inspiracional[94].

Como puede observarse, las vulnerabilidades que experimentan cada una de estas poblaciones se van sumando y se expresan también en el espacio digital. El primer nivel de desigualdad se plasma en el acceso diferenciado a Internet, mismo que se va articulando con otras limitantes. Un contrapeso frente a las brechas en el acceso a Internet y a la información es la visibilidad y la producción de líderes de opinión a favor de estas poblaciones. Sin embargo, en el caso de las redes sociodigitales, las personas especialistas técnicas entrevistadas refieren que es imposible que una red presente a las personas usuarias información de todos sus contactos, siendo que el criterio empleado para decidir qué información se presenta y cuál no, responde a fines comerciales. Así, el potencial comercial influye en la difusión de unas voces por encima de otras.

Seguridad personal, salud mental y agencia

Una de las afectaciones que refirieron con mayor énfasis las personas entrevistadas fue la relativa a la salud mental derivada de las amenazas, la estigmatización y el desgaste. Respecto a las amenazas, muchas de ellas provienen de cuentas privadas o anónimas. Si bien, no siempre coartan directamente la libertad de expresión, sí tienen un efecto inhibidor en la manera de expresarse o comunicarse, además del impacto a nivel de salud (cuyo nivel máximo es el suicidio) cuando las personas están constantemente siendo amedrentadas, acosadas e hipervigiladas sistemáticamente en redes sociodigitales. Al respecto, la población LGBT+ y algunas mujeres son objetivo de muchas amenazas de muerte y violencia sexual; además, se viralizan sus rostros y se comparten sus datos personales.

También, algunas personas abren perfiles fingiendo ser mujeres para desprestigiarlas a partir de la sanción social sobre su libertad sexual. Aunado a lo anterior, las políticas de privacidad de ciertas redes impiden -a menos que exista una orden judicial de por medio- solicitar información de las personas agresoras, por lo que no siempre es posible señalar directamente la fuente de la agresión o, en el caso que sea posible solicitar dicha información,

94 Término acuñado por la activista Stella Young en 2012 para señalar el uso de las personas con discapacidad para producir mensajes inspiradores basados en su discapacidad (Wikipedia, s.f.). Wikipedia. (Sin fecha). Porno inspiracional. Disponible en https://es.wikipedia.org/wiki/Porno_inspiracional

los mecanismos para hacerlo son sumamente complejos. Por su parte, en los talleres las personas periodistas expresaron sentir frustración y tristeza por las limitaciones que enfrentan para realizar su profesión. Estas emociones fueron catalogadas como daños emocionales que se enlazan con daños en lo laboral.

Una de las formas en que estos grupos enfrentan las amenazas a su libertad de expresión en espacios digitales es la *autocensura*. Esta consiste en la decisión de las personas usuarias de retirar algún contenido o dar de baja –temporal o definitivamente– su perfil-página de alguna red. Al respecto, 87% de las personas encuestadas reportaron que se han limitado en al menos una ocasión a publicar algún contenido por miedo a la reacción que pudieran desatar; además, 94% refirió conocer a alguien que también se ha auto restringido el hacer alguna publicación por la misma razón. Para las personas entrevistadas, especialmente para aquellas que integran las poblaciones de mujeres y LGBT+, la autocensura significa una medida de autocuidado y protección, sin embargo, también representa invisibilización, silenciamiento doloroso, ruptura comunicacional y límites a su socialización. La autocensura es una decisión que las personas entrevistadas refirieron como una medida de autocuidado, cuando sostener su presencia en redes sociodigitales es insostenible debido al acoso y/o a la exposición repetida de discursos de odio. Otras estrategias relacionadas con la autocensura son, por ejemplo, configurar sus perfiles de modo privado (para el caso de Instagram) y restringir quién puede comentar sus tuits.

La autocensura fue un tema presente en los diversos espacios de interlocución con quienes fueron informantes (entrevistas, talleres y encuesta) ya que encierra ciertas tensiones. Por ejemplo, la paradoja de la seguridad *versus* visibilidad: es importante para los grupos en condición de vulnerabilidad hacerse presentes en plataformas digitales, pero al mismo tiempo su presencia les implica riesgos potenciales a los que se exponen justo por hacerse visibles. Limitar su voz y presencia a través de la autocensura les protege de dichos riesgos, pero les hace perder seguidores(as), lo cual debilita sus redes de apoyo (reales o potenciales). En ese sentido, como refirió una de las personas asistentes a talleres, la libertad de expresión no es un derecho absoluto, sino que cobra distintos matices en contextos violentos.

Otra paradoja que se discutió en los talleres, especialmente el realizado en la ciudad de Mérida, es que cuando se responde a agresiones se suelen infringir normas comunitarias, se produce el mensaje de que defenderse es algo que se castiga, inhibiendo así que las personas lo hagan o, en su caso, se revictimiza a través del silenciamiento. Este tipo de censura que ejercen las plataformas refuerza la idea de que las minorías no sólo no tienen argumentos válidos, sino que no existen espacios de enunciación para ellas.

Opuesto a lo que ocurre con la población LGBTQ+, la visibilidad en redes sociodigitales beneficia a la población indígena para hacer pública la lucha por la defensa de la tierra. En Yucatán, por ejemplo, la visibilidad ha abonado incluso a la seguridad de las personas defensoras de la tierra pues en caso de ser agredidas, su visibilidad permite que autoridades y medios de comunicación otorguen importancia a la agresión en caso de que esta ocurra. No obstante, es importante señalar que no sucede así en todos los casos, como es las comunidades que se oponen a la construcción del Tren Maya, el cual se trata de un megaproyecto que cuenta con el respaldo político del Poder Ejecutivo Federal en turno. En ese sentido, se deduce que el efecto positivo de la visibilización para esta población en espacios digitales varía en función de factores tales como el nivel de poder del actor que respalda el proyecto al que se oponen.

En el caso de la población migrante, dado su carácter de “ilegal”, no suelen exponerse en redes sociodigitales como población migrante. El modo en que agregan contactos a su red es más selectivo (más privado), así, aunque en círculos limitados, pero pueden permanecer dentro de lo digital. Por su parte, el seguimiento digital que se hace de las caravanas migrantes, también sirve de protección para las mismas, pues disminuye las probabilidades de que sean víctimas de grupos delincuenciales.

Respecto a cómo formular las publicaciones en redes sociodigitales para reducir las probabilidades de ser denunciadas o acosadas, una de las personas entrevistadas activista por los derechos de las mujeres refirió que han notado que los comunicados contestatarios producen muchos “me enoja” y “me divierte”, mientras que los comunicados escritos en un tono “amigable” reciben más “me encanta” y “me gusta”. Estas sanciones o premios, respectivamente, muestran una interpretación negativa digna de ser rechazada (“me enoja”) o burlada (“me divierte”) cuando

provienen de mujeres ya que, por lo general, existe una lectura de que se trata de una agresión. Al respecto, la representación social diferenciada entre mujeres y hombres como sujetos legitimados para el uso de la violencia modela la práctica de sanción cuando son ellas quienes la ejercen (o parecen ejercerla) y, en particular, cuando son mujeres que de una u otra forma no se ajustan al mandato social de feminidad. Los significados asociados a lo femenino y a lo masculino, las representaciones diferenciadas de cada género y la internalización que las personas hacen de estas diferencias muestran al orden de género como un aparato articulado que justifica y sostiene la desaprobación de, en este caso, "lo contestatario" en las mujeres. Estas sanciones vinculadas a la expresión de emociones dan cuenta de una distribución desigual de poder (Hochschild, 1975).

En cuanto a la seguridad de las redes, no hay un consenso sobre cuál es la más segura, pero sí hay acuerdo en que Facebook y Twitter son las que generan más estrés por la vigilancia y excesiva circulación de "discursos de odio". Estos discursos se pueden denunciar, pero también existen las contra-denuncias (denunciar a quien denuncia), muchas de las cuales se fundamentan en que alguna publicación "no gusta", en lugar de por violentar a alguien. Todo lo anterior conlleva a que las redes sociodigitales se tornan en espacios donde la libre expresión se pone en duda y las personas sostienen una actitud a la defensiva.

Conclusiones y recomendaciones

Las plataformas digitales juegan un papel importante en el ejercicio de la libertad de expresión y el acceso a la información. En México, las personas usuarias pasan (en promedio) 4 hrs con 8 minutos[95] en redes sociales diariamente, espacio digital donde se expresan e informan. Las personas en situación de vulnerabilidad que presentamos en este estudio padecen de las limitantes de las brechas digitales y de diversas *barreras técnicas* tanto para el ejercicio de sus derechos digitales como para el acceso a otros derechos.

95 Estudio de Hábitos de 2020 de los usuarios de Internet en México, 2020.

Las redes sociodigitales son espacios que pretenden ser seguros y accesibles, que tienen como objetivo la creación de una comunidad en la cual todas las personas puedan crear y compartir contenido; sin embargo, la investigación desarrollada nos muestra otra cara totalmente diferente, espacios con comentarios racistas, nocivos, violentos y estigmatizantes, lo cual resulta ser un reflejo de la realidad que se vive en México. En este sentido, la moderación de contenido, juega un papel muy importante en la aplicación de criterios (normas comunitarias) de cada red social. A pesar de la diversidad de la comunidad que participa activamente en las plataformas digitales, las normas comunitarias son reglas generales y abstractas que en muchas ocasiones no consideran los casos y contextos específicos.

Es importante reconocer que existen factores que complejizan la moderación de contenido como: la gran cantidad de personas usuarias (en algunos casos); las diferencias del lenguaje y contexto; la clasificación del tipo de contenido; y la estructura de las plataformas digitales y su modelo de negocio. Por tal razón, algunas plataformas utilizan la *moderación automatizada de contenido* como herramienta de filtrado *ex-ante*.

Desde las voces de las personas participantes en este proyecto, se considera que:

A nivel **institucional**, las asociaciones civiles son el agente que más gestiona en pro de las poblaciones en situación de vulnerabilidad, incluido el dotarlas de recursos y conectividad a Internet; en sentido opuesto, el Estado se deslinda de sus responsabilidades. En función de lo anterior, resulta indispensable, en primer lugar, tener como punto de partida que los derechos están vinculados entre sí, de manera que si, por ejemplo, no se atiende la condición de pobreza en que vive la población indígena, será imposible garantizar otros derechos, incluidos los digitales. En ese sentido, se requieren fortalecer los marcos de derechos humanos fuera de Internet, así como de aquellos que cada vez más dependen de la posibilidad de conectividad, ya que la violencia que ocurre en el ámbito digital está conectada con la que ocurre fuera y le antecede, tal es el caso de la corrupción estructural de las instituciones que ha contribuido a normalizar la violencia en general. Aunado a lo anterior, dentro de los talleres se destacó la necesidad de garantizar el acceso universal a Internet a la par de brindar alfabetización digital,

capacitación en temas de seguridad digital y funcionamiento de las redes sociodigitales, pues de otro modo, la libertad de expresión en el espacio digital es impensable. Así mismo, es importante que el Estado promueva alianzas institucionales y participe en la producción de un piso mínimo para la regulación en Internet. Respecto a este punto, la duda que prevalece es si este agente está capacitado para participar en la regulación o, si por el contrario, su intervención resultaría perjudicial.

En lo que respecta a la estructura de las **redes sociodigitales**, las personas entrevistadas y asistentes a los talleres refieren que debería existir una propia regulación algorítmica que permita saber si tanto la información compartida como los perfiles son verídicos. De igual forma, se observa como necesario que las políticas y criterios empleados para bloquear una cuenta sean realmente claros y diseñados no para castigar, si no para garantizar la interacción no violenta; esta última refiere también a cómo las plataformas explican a la persona usuaria los motivos por los que su cuenta es bloqueada, dejando espacio para la interlocución. Bajo el hecho de que las redes sociodigitales amplifican unas voces y silencian otras, la recomendación no es silenciar a las que ya cuentan con presencia, sino reestructurar la arquitectura de las redes para equilibrar la presencia de todas; además, se sugiere que estas redes cuenten con un(a) ombudsperson capacitado(a) y sensible a las distintas vulnerabilidades. Resulta evidente que las violencias y vulneraciones de derechos que se expresan en los espacios digitales requieren intervenciones complejas tanto en lo *online* como en lo *offline*. La forma en que opera la moderación automatizada de contenido es, en parte, reflejo de las estructuras de violencia y desigualdad que preceden la existencia de Internet y, hoy día, es claro que obedece a intereses capitalistas. Al respecto, una de las personas entrevistadas comentó que muchas de las personas que participan en la programación algorítmica apoyan individualmente a los grupos vulnerables que hemos mencionado, el problema es que los intereses de las empresas para las que trabajan no se rigen con sus mismos principios.

Referente a las **personas y asociaciones defensoras de derechos humanos**, conviene recordar que frente a las limitantes que impone la moderación automatizada y de la escasa participación del Estado, estos agentes cumplen un rol fundamental para informar a las poblaciones que hemos hecho mención. Sin embargo, para el caso

de la población migrante resalta que es indispensable desarrollar información inmediata y simple, pero que conserve calidad y veracidad; ambas necesidades obedecen a que la desinformación *esperanzadora*[96] se propaga con rapidez. Sobre este punto, cabe aclarar que la propagación veloz no obedece solo a la acción de las personas, sino también a las dinámicas propias de cada plataforma, en ese sentido, se trata de una responsabilidad compartida. Como medida para contrarrestar el esparcimiento de información falsa, se considera importante fortalecer los equipos de comunicación de las asociaciones civiles –y también los institucionales–. Así mismo, se señaló que estos agentes deben ser cuidadosos de no promover información sensacionalista, pero parcialmente falsa, solo por alcanzar visibilidad. Lo anterior es muy peligroso porque los grupos antiderechos suelen manejar información para refutar en estos casos, dando pie así, indirectamente, a la producción de discursos de desprestigio y odio. De igual forma, se recomienda desarrollar indicadores para medir riesgos, así como estrategias de autocuidado, procuración de la salud mental y establecimiento de límites individuales dentro de su labor como activistas.

Finalmente, las personas entrevistadas compartieron algunas recomendaciones que apelan a la *población usuaria de redes sociodigitales* en general. Por una parte, instan a realizar un ejercicio de reflexividad para cuestionar las motivaciones para replicar violencias en el espacio digital desde una perspectiva crítica y con capacidad de escucha-diálogo. Por ello, es importante construir ciudadanía digital y una actitud abierta a la diversidad de opinión y búsqueda de información dentro del marco de los derechos humanos. Otras de las acciones sugeridas es que las personas usuarias deberían realizar un análisis contextual de todo contenido al que se acceden en Internet. Así mismo, se considera importante visibilizar las prácticas de vigilancia y censura y crear redes de apoyo formales entre activistas para compartir saberes y medidas de cuidado.

Desafortunadamente, por lo general las redes sociodigitales que respetan la diversidad, no han logrado posicionarse por no ser rentables en términos comerciales, algunas ponen en riesgo los datos personales de las personas usuarias. De forma específica, en lo que respecta a las personas con discapacidad se recomienda hacerlas partícipes dentro de las redes como sujetos de derechos y no como “un(a) otro(a) vulnerable”; hacerlo así impide reconocer sus necesidades y derechos. Si se quiere incluir a las diversidades, se

debe hacer desde un marco plural, participativo y horizontal. En este punto, se subraya que todas las personas deberían tener el mismo derecho de tener una cuenta en redes sociodigitales y mostrar ahí su vida como lo deseen, libres de tabú y prejuicios por parte de los grupos dominantes.

En relación con todo lo anterior, como equipo de trabajo establecemos las siguientes recomendaciones desde una perspectiva de derechos humanos y de gobernanza de Internet aplicado a la moderación de contenido:

Primero.- La *moderación automatizada* de contenido tiene problemas como la falta de capacidad de interpretar el contexto antes de la toma de decisión. Esta situación es un riesgo para la libertad de expresión y el acceso a la información, por lo cual resulta relevante y necesario analizar el contexto al momento de llevar a cabo un proceso de moderación. Diversos casos los podemos encontrar en las recomendaciones del Consejo Asesor de Contenido de Facebook[97] desde el 2020, organismo que garantiza el ejercicio de la libertad de expresión mediante un proceso independiente. Es por ello, que resulta apremiante una discusión profunda con especialistas de las áreas relevantes que nos permitan identificar las problemáticas lógicas, lingüísticas, informativas, semióticas, históricas, sociales, culturales y jurídicas del contexto en los discursos y su moderación. Es importante mencionar que en las redes sociodigitales, no únicamente la moderación de contenido afecta la libertad de expresión de las personas en situación de vulnerabilidad, sino que encontramos otros factores que debilitan este derecho como: el modelo de negocio y la arquitectura de las plataformas digitales, la brecha digital y la dinámica social.

Segundo.- Podemos mencionar que resulta relevante *impulsar el diálogo entre las plataformas digitales*[98]. Las situaciones que experimentan las personas en situación de vulnerabilidad, pueden ser tratadas de forma similar por las plataformas digitales. Es decir, las experiencias pueden ser tomadas en consideración como un aprendizaje, no únicamente interno sino colectivo, y que puede servir para la toma de decisiones y mecanismos externos de moderación de contenido.

Tercero.- La creación de instrumentos de gobernanza de Internet vinculados con la moderación de contenidos de las plataformas

96 De acuerdo con las personas entrevistadas, este tipo de información incentiva a las personas a continuar el tránsito migratorio al prometerles algún beneficio, ayuda rápida o garantizar la seguridad mientras estén en movilidad. Al tratarse de información alentadora, suele cuestionarse poco su veracidad y se la comparte sin previa verificación.

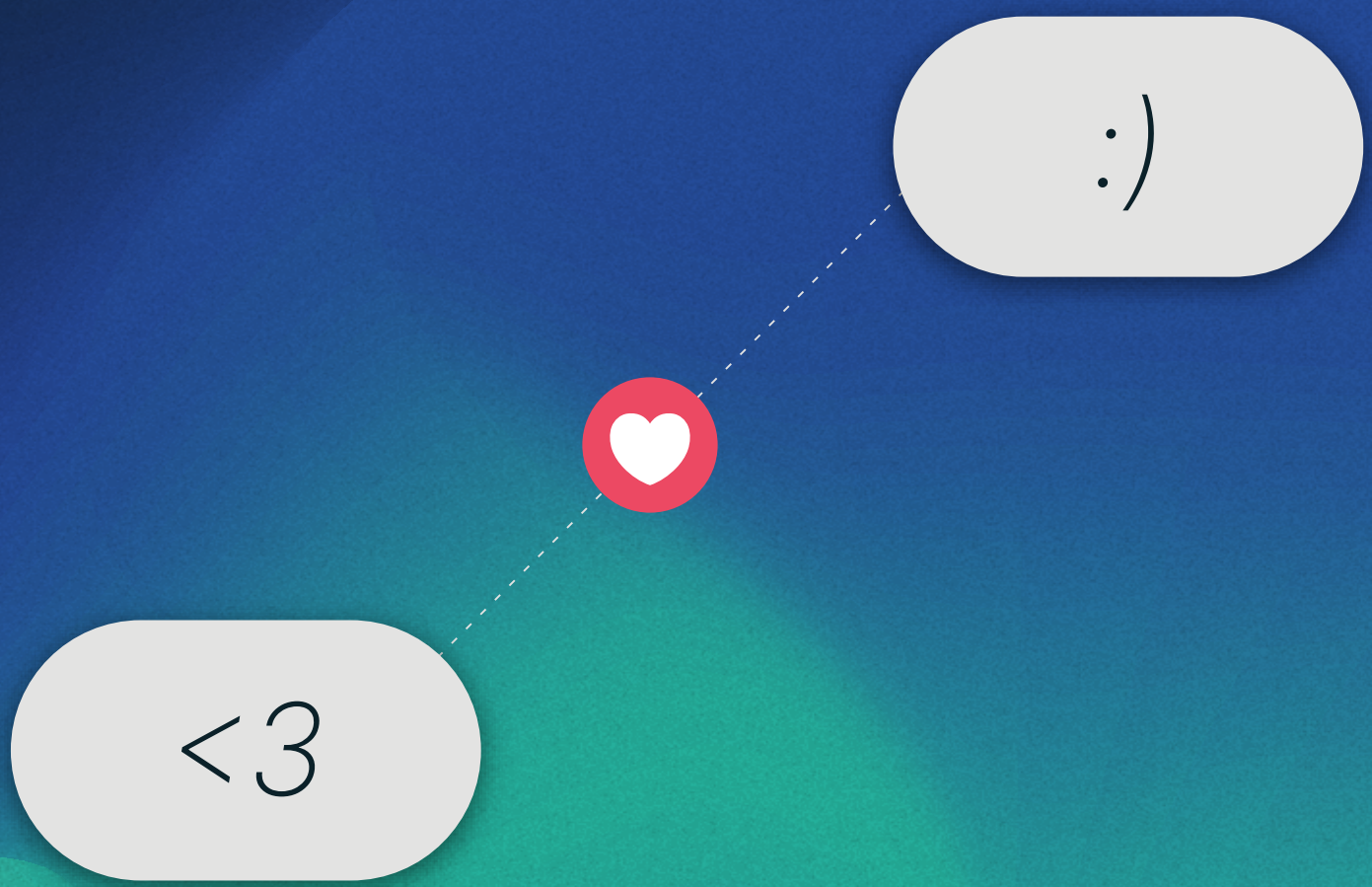
97 https://oversightboard.com/

98 Lanza, Edison y Matías, Jackson, Moderación de Contenidos y Mecanismos de Autorregulación. “El Oversight Board” de Facebook y sus implicancias para América Latina, El Diálogo, 2021, p. 24.

digitales, deben llevarse a cabo mediante un trabajo colaborativo (multistakeholder). Las perspectivas de diferentes actores son necesarias en los mecanismos de creación, razón por la cual resulta trascendental impulsar espacios de discusión multisectoriales, así como la creación de instituciones que cuenten con representación plural no solamente de los tomadores de decisiones, sino de los sujetos de dichas decisiones. El caso más relevante son los grupos históricamente discriminados que presentamos en este estudio y las comunidades en situación de vulnerabilidad.

Cuarto.- Durante la pandemia pudimos observar cómo Meta agregó etiquetas a contenidos sobre tratamientos sin pruebas científicas. En este sentido, pudimos observar como la citada red social trabajó cerca con autoridades de la salud alrededor del mundo para conectar a las personas con información segura y actualizada sobre prevención y vacunas. Desde que la OMS declaró al COVID-19 como una emergencia de salud pública, Meta ha colaborado para que las personas usuarias tengan información confiable a miles de millones de personas. Esto es así ya que las consecuencias de una desinformación acerca de este rubro pueden poner en riesgo la salud no es en estricto sentido mecanismo de moderación que, como en el ejemplo citado, funge como herramienta para combatir la desinformación. El contenido que es relevante para activistas, personas defensoras de derechos humanos y periodistas, puede ser etiquetado para crear espacios seguros.

Quinto.- En lo que respecta a la autocensura, se observó que se la considera una medida de autocuidado digital en contextos de violencia y gravita entre la agencia de las personas usuarias de redes sociodigitales y la auto restricción de la libertad de expresión que aun cuando se gesta fuera del agente lo lleva a ser quien la ponga en marcha. Sin embargo, cuando las personas buscan cómo convivir y sortear las restricciones de la moderación de contenido y violencia en espacios digitales, no debe perderse de vista que son efectos de la vulneración de derechos, al tiempo que nos muestra que los derechos no están garantizados de manera permanente, por lo que necesitan estarse defendiendo constantemente.


:)
<3

Fuentes de Información Consultadas

Bibliográficas

Artículo 19. (7 de agosto de 2020). ¿Qué es el discurso de odio?

Artículo 19. (2020). COVID. Article 19 Oficina para México y Centroamérica. pp. 25.

Artículo 19. (2020). #LibertadNoDisponible. Censura y remoción de contenido en Internet Caso: México.

Artículo 19. (2022). Negación. Informe anual 2021.

Asociación de Internet MX. 18° Estudio sobre los hábitos de personas usuarias de Internet en México. Mayo 2022.

Asociación de Internet MX. Estudio de Hábitos de 2020 de los usuarios de Internet en México, 2020.

Berrío, C., Marín, P., Ferreira, E. y Chagas, E. (2017). Desafíos de la inclusión digital: antecedentes, problemáticas y medición de la brecha digital de género. Psicología, Conocimiento y Sociedad.

Calderón, F., y Castells, M. (2019). La nueva América Latina. Fondo de Cultura Económica.

Cantoral, Karla. Daño moral en redes sociales: su tratamiento procesal en el derecho comparado. Rev. IUS [online]. 2020, vol.14, n.46 pp.163-182.

Cloudfare. (Sin fecha). ¿Qué es la nube?

Committee to Protect Journalists. (Sin fecha). Los 10 países con la mayor censura.

Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), Resultados de pobreza en México 2020 a nivel nacional y por entidades federativas, 2020.

Courtis, citado por González, M. y Parra, O. (2008). Concepciones y cláusulas de igualdad en la jurisprudencia de la Corte Interamericana. A propósito del Caso Apitz. *Revista IIDH*. pp. 135.

Cruz, D. «Afecto, cuerpo e identidad: reflexiones encarnadas en la investigación feminista». En Rabasa y Guerrero Mc Manus, Afecto, cuerpo e identidad. Reflexiones encarnadas en la investigación feminista, 238., editado por Alba Pons Rabasa y Siobhan Guerrero Mc Manus. Serie Estudios jurídicos, núm. 330. Ciudad de México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas, 2018.

Escamilla, Sandra yPantin, Laurence. “La Justicia Digital En México: El Saldo A Un Año Del Inicio De La Pandemia - México Evalúa”. México Evalúa. 2021.

Flew, Terry, Martin, Fiona, & Suzor, Nicolas. Internet regulation as media policy: Rethinking the question of digital communication platform governance. Journal of Digital Media and Policy, 10(1), pp. 33-50.2019.

Galtung, J. (1990). Cultural Violence. *Journal of Peace Research*, 27(3), 291–305.

Giovanni de Gregorio, en Luca Belli, Nicolo Zingales y Yasmin Curzi Terms, “Glossary Law and Policy”, 2021.

Gómez, D., Alvarado, R., Martínez, M. y Díaz, C. (2018). La brecha digital: una revisión conceptual y aportaciones metodológicas para su estudio en México. Entreciencias: diálogos en la sociedad del conocimiento.

Greenspan, R. y Tenbarge, K. (26 de septiembre de 2020). YouTuber’s channels and videos are being mistakenly deleted for debunking COVID-19 conspiracy theories. *Insider*.

Grimmelmann, James, “The Virtues of Moderation”, Yale Journal of Law and Technology, EEUU, pp.42.

Guillén López, Tonatiuh. México, nación transterritorial. El desafío del siglo XXI , Universidad Nacional Autónoma de México, Programa Universitario de Estudios del Desarrollo, 2021.

Gutiérrez, A. (2016). Etnografía móvil: una posibilidad metodológica para el análisis de las identidades de género en Facebook. Revista interdisciplinaria de estudios de género de El Colegio de México.

Haimson, O., Delmonaco, D., Nie, P., y Wegner, A. (2021). Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. *Proceedings of the ACM onHuman-Computer Interaction*, 5(CSCW2), 1-35.

Hochschild, A. (1975). The sociology of feeling and emotion: Selected possibilities. *Sociological Inquiry*, 45(2-3), 280-307.

Instituto Nacional de Estadística y Geografía (INEGI), Encuesta Nacional de Cultura Cívica (ENCUCI) 2020.

Instituto Nacional de Estadística y Geografía . Encuesta Nacional de Calidad e Impacto Gubernamental 2021.

Instituto Nacional de Geografia y Estadística, Encuesta Nacional sobre Discriminación (ENADIS) 2017.

Instituto Nacional de Geografia y Estadística, Encuesta Nacional sobre Discriminación (ENADIS) 2017.

Instituto Nacional de Geografía y Estadística, Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2020.

Instituto Nacional de Geografía y Estadística. Censo de Población y Vivienda 2020. D

Lanza, Edison y Matías, Jackson, Moderación de Contenidos y Mecanismos de Autorregulación. “El Oversight Board” de Facebook y sus implicancias para América Latina, El Diálogo, 2021, pp.

Lanza, Edison y Matías, Jackson, Moderación de Contenidos y Mecanismos de Autorregulación. “El Oversight Board” de Facebook y sus implicancias para América Latina, El Diálogo, 2021.

Llansó, E. J. (2020). No amount of “AI” in content moderation will solve filtering’s prior-restraint problem. *Big Data & Society*, 7(1), 2053951720920686.

Maldonado, J. (2017, marzo 18). Repensar la práctica del cuidado en el contexto del síndrome de Down. Debate Feminista, 53.

Mendoza, J. y Caldera, J (2014). Umbrales para la determinación de la brecha digital: comparativa entre regiones desarrolladas. Transinformação, 26.

Mohanty, Chandra Talpade. Feminism without borders: decolonizing theory, practicing solidarity. Durham:Duke University Press. 2003.

Montezuma, Luis Alberto. The case for a hybrid model on data protection/privacy. International Association of Privacy Professionals, May 6, 2020.

Oben Uru, F. The Dark Side of Digitalization: Digital Mobbing. In F. Ozsungur (Ed.), Handbook of research on digital violence and discrimination studies. Business Science Reference. 2022.

Padrón Innamorato, Mauricio, Disponibilidad y acceso a la tecnología como una aproximación para el estudio del fenómeno de acceso a la información y su relación con la pobreza en México, en Luna Pla, Issa, and Universidad Nacional Autónoma de México. 2014. Estudios aplicados sobre la libertad de expresión y el derecho a la información. 1° ed. México D.F.: Universidad Nacional Autónoma de México Instituto de Investigaciones Jurídicas.

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., y Tacchi, J. (2019). Etnografía digital. Ediciones Morata.

Primer estudio sobre los hábitos de Médicos de Internet en México, abril 2021.

R3D. Gobierno Espía. Vigilancia sistemática a periodistas y defensores de derechos humanos en México. 2017.

Revilla, M. Participación política: lo individual y lo colectivo en el juego democrático. En Benedicto, J. y Morán, M. (eds.). Sociedad y política. Temas de sociología política. Alianza editorial. Segunda reimpresión. Madrid; 2017.

Salazar Ugarte, Pedro y Gutiérrez Rivas, Rodrigo, El derecho a la libertad de expresión frente al derecho a la no discriminación. Tensiones, relaciones e implicaciones, México, UNAM, Instituto de Investigaciones Jurídicas-Consejo para Prevenir la Discriminación, 2008.

Scholte, J. A. (2017). Polycentrism and democracy in internet governance. *The net and the nation state: Multidisciplinary perspectives on Internet governance*, 165-184.

Secretaría de Gobernación. (2021, abril). Acuerdo por el que se da a conocer el Programa de Conectividad en Sitios Públicos 2020-2021 de la Secretaría de Comunicaciones y Transportes..

Solum, Lawrence, “Models of Internet governance”, Bygrave, Lee A. y Bing Jon (eds.), Internet Governance. Infrastructure and Institutions, New York, Oxford University Press, 2009.

Unión Internacional de Telecomunicaciones, Estadísticas, 1996.

UN Human Rights. (8 de marzo de 2017). UN experts urge States and companies to address online gender-based abuse but warn against censorship.

Instrumentos Internacionales y Normatividad

CIDH. Caso Gonzales Lluy y otros Vs. Ecuador. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 1 de septiembre de 2015. Serie C No. 298.

SCJN, Protocolo para juzgar con perspectiva de discapacidad, abril de 2022.

CEDAW, Convención sobre la Eliminación de Todas las Formas de Discriminación Contra la Mujer 1979.

CIDH, Declaración de principios sobre libertad de expresión. 2020.

CIDH, Relatoria Especial para la Libertad de Expresión, Libertad de expresión e Internet, OEA/Ser.L/V/II CIDH/RELE/INF.11/13, 31 diciembre 2013

CIDH. Condición jurídica y derechos de los migrantes indocumentados. Opinión Consultiva OC-18/03 de 17 de septiembre de 2003. Serie A No. 18.

Comisión Interamericana de Derechos Humanos. Marco jurídico interamericano sobre el derecho a la libertad de expresión. New York; CIDH; Asdi: OEA ;2010.

ONU, Informe de la Alta Comisionada de las Naciones Unidas para los Derechos Humanos acerca de los talleres de expertos sobre la prohibición de la incitación al odio nacional, racial o religioso. 2013.

Organización de las Naciones Unidas, Informe del Relator Especial sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental, A/HRC/41/34, 12 de abril de 2019.

Relator Especial de las Naciones Unidas (ONU) sobre la Promoción y Protección del derecho a la Libertad de Opinión y de Expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), también en UNESCO, Global toolkit for judicial actors. International legal standards on freedom of expression, access to information and safety of journalists. 2021.

Relator Especial de las Naciones Unidas (ONU) sobre la Promoción y Protección del derecho a la Libertad de Opinión y de Expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión, y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). 1 de junio de 2011, Principio 3. C.

Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión, y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). 1 de junio de 2011.

Relatoría Especial para la Libertad de Expresión, Comisión Interamericana de Derechos Humanos, OEA, Open Society Foundations. Libertad de expresión e Internet, diciembre 2013.

Diario Oficial de la Federación. Decreto por el que se declara(n) reformadas y adicionadas diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, relativos al Poder Judicial de la Federación. 11 de marzo de 2020 2021.

Anexos

Anexo 1. Equipo de trabajo

Para el desarrollo del proyecto, trabajamos con un equipo multidisciplinario proveniente de cuatro instituciones:

Sobre la oficina de Artículo 19 México y Centroamérica

Artículo 19 es una organización no gubernamental independiente que promueve y defiende la aplicación progresiva de los derechos de la libertad de expresión y el acceso a la información para todas las personas, conforme a los más altos estándares internacionales de derechos humanos, de tal manera que se contribuya al fortalecimiento de la democracia. Para cumplir su misión, Artículo 19 tiene las siguientes tareas principales: exigir el derecho a difundir información y opiniones en todos los medios de comunicación, investigar las amenazas y los avisos, documentar las violaciones a los derechos a la libertad de expresión, proporcionar apoyo a las personas cuyos derechos han sido violados y ayudar a diseñar políticas públicas en su área de acción. En este sentido, Artículo 19 prevé una región donde todas las personas puedan expresarse en un entorno de libertad, seguridad e igualdad, y ejercer su derecho a acceder a la información; de tal modo que se facilite la incorporación de la sociedad a la toma de decisiones informadas sobre lo que implica por sí mismas y en su entorno, para la realización plena de otros derechos individuales.

Artículo 19 trabaja para vincular la promoción de políticas públicas, el acompañamiento a procesos locales de organizaciones y el ejercicio de los derechos humanos en varias entidades en México y Centroamérica. Artículo 19 promueve el reconocimiento y la protección de los derechos humanos en entornos digitales, particularmente el derecho a la libertad de expresión e información para evitar el establecimiento y la práctica de mecanismos de censura en Internet o medidas que obstaculicen su ejercicio a través de la legislación, las políticas públicas, los tratados internacionales, las decisiones judiciales o administrativas, o las iniciativas privadas. Artículo 19 trabaja para garantizar las condiciones adecuadas para

que las personas, los medios de comunicación y los/las periodistas ejerzan sus derechos a la libertad de expresión e información, a la privacidad, al acceso a Internet sin discriminación y a cualquier otro derecho pertinente en un ecosistema digital. Del equipo de Artículo 19, se han sumado a este proyecto Priscilla Ruíz y Vladimir Cortés.

Cultivando Género A.C.

Cultivando Género A.C. es un grupo de mujeres que hemos formado una Asociación Civil cuyo objeto es promover la igualdad y derechos de mujeres, niñas y niños, a través de la educación para la paz y derechos humanos. Ofrece servicios especializados, estratégicos e integrales de prevención, atención, acompañamiento y asesoría en temas vinculados a los derechos humanos y la perspectiva de género, dirigida a organizaciones públicas, privadas y sociales, con el fin de incidir en la transformación positiva de sus ambientes laborales, educativos y sociales.

En Cultivando Género trabajamos en 4 ejes estratégicos en los que se enmarca esta investigación: 1) Capacitación y sensibilización en materia de género; 2) Violencia digital; 3) Cultura organizacional para la prevención de violencia en ámbitos laborales; y 4) Derecho al cuidado como derecho humano. Participaron directamente en el desarrollo de este proyecto Angie Contreras (feminista, comunicadora social) y Wina Rosas (socióloga feminista).

Colectivo por la Protección de Todas las Familias en Yucatán

El Colectivo PTF Yuc es una organización dedicada a la concientización a través de campañas educativas, de cabildeo y litigio estratégico a favor de los derechos humanos de la comunidad LGBTQ+ y sus familias en Yucatán. Fundado el 6 de marzo de 2019 en respuesta a la urgencia de organizar un frente estratégico para el cabildeo y litigio a favor del matrimonio igualitario en Yucatán. Integrado por un grupo núcleo de seis amistades activistas y por más de dos mil seiscientas personas activas en el grupo de Facebook.

El Colectivo PTF Yuc emprendió cuatro demandas de amparo en contra de la Legislatura LXII del Congreso de Yucatán, tres de ellos por el formato de la votación y uno por violar, con las votaciones por cédula (secretas) en contra del matrimonio igualitario. El 19 de agosto de 2021, la Primera Sala de la SCJN falló por unanimidad a favor de los amparos 25/2021 y 27/2021, instruyendo al Congreso de Yucatán a reparar el procedimiento de votación de la reforma constitucional para permitir el matrimonio igualitario. Del Colectivo participan directamente: Alex Orué y César Briceño.

Instituto de Investigaciones Jurídicas de la UNAM. Línea de Investigación en Derecho e Inteligencia Artificial (LIDIA y el Laboratorio Nacional Diversidades (LND)

El Instituto de Investigaciones Jurídicas de la Universidad Nacional Autónoma de México tiene como uno de sus objetivos principales el desarrollo de investigaciones sobre las problemáticas jurídicas nacionales, así como la formación de profesionales aptos para la atención de dichas problemáticas. Adicionalmente, se le encarga la difusión de los diferentes acercamientos para entender y resolver estos retos entre la población para reducir brechas de conocimiento y potenciar su discusión. Muchos de los problemas hoy trascienden las fronteras nacionales y conjugan un carácter novedoso para la mayoría de nosotros. Las tecnologías y con ellas Internet nos enfrentan a viejas preguntas en nuevos contextos. El reto de analizar problemas tan complejos es intentar abordar únicamente los puntos necesarios en la explicación y no perecer en el intento de ordenarlos de una manera comprensible.

A través de su la Línea de Investigación en Derecho e Inteligencia Artificial del Instituto de Investigaciones Jurídicas, ha llevado a cabo diversos ejercicios de reflexión, discusión y análisis para conocer, problematizar y, quizá, comenzar a entender la problemática y efectos jurídicos derivados de la implementación de las nuevas tecnologías. Entre ellos, el *control de la conducta en la red* y en especial la *moderación de contenidos* han concentrado recursos y acuerdos sobre su imperiosa necesidad de análisis. Desde este proyecto y otros previos y paralelos, hemos tenido la oportunidad de escuchar a una extensa gama de expertos en la materia del sector privado, de

la sociedad civil organizada, académicos, defensores y víctimas de abuso en línea. Participaron de manera directa en este proyecto: Pablo Pruneda Gross, Jesús González Mejía y Francisco Chan Chan.

Adicionalmente, con el apoyo del Laboratorio Nacional Diversidades (LND) es una unidad de investigación alojada en el Instituto de Investigaciones Jurídicas de la UNAM para el desarrollo científico y la innovación en temas de diversidad, género e inclusión. Es un espacio plural e incluyente que promueve la investigación, divulgación, vinculación e incidencia en torno a las diversidades en distintos grupos sociales, y de manera particular en dos poblaciones: las personas migrantes y las juventudes.

Desde un enfoque interseccional e interdisciplinario que incluye al derecho, la sociología, la antropología, la demografía, la ciencia política, los estudios culturales y de género, el LND produce conocimiento e incidencia social sobre las condiciones que generan obstáculos y avances para la inclusión de las diversidades en distintos contextos como pueden ser la comunidad universitaria y las comunidades migrantes. Colaboraron con este proyecto Alethia Fernandez de la Reguera Ahedo y Adriana Figueroa Muñoz Ledo.

Anexo 2. Formato de encuesta

Objetivo: Identificar las experiencias relacionadas con la libertad de expresión y moderación de contenido en plataformas de redes sociales.

Gracias por participar en el taller "Inteligencia Artificial, remoción de contenido y libertad de expresión en la era digital" el cual busca generar, intercambiar y documentar las experiencias de algunas poblaciones de atención prioritaria sobre los principales problemas que enfrentan en el ejercicio de sus derechos a la libertad de expresión y acceso a la información en plataformas digitales.

Bienvenide,

Antes de contestar la encuesta es importante hacerte saber que:

* *La recolección y tratamiento de la información será realizada exclusivamente por el equipo de Artículo 19.*
* *Cualquier duda puedes escribirnos a priscilla@article19.org*
* *No recopilamos correos electrónicos o datos personales.*
* *En el momento que lo consideres puedes detener la encuesta.*
* *La información es anónima.*

Gracias por tu apoyo.

a) Datos Generales:

1. Edad _____ (años cumplidos)

2. Indique su género:

◯ Mujer ◯ Hombre ◯ Mujer Transgénero ◯ Hombre Transgénero
◯ No Binario ◯ Prefiero no decirlo

3. De acuerdo con tu orientación sexual, te identificas como:

◯ Lesbiana ◯ Homosexual ◯ Bisexual
◯ Heterosexual ◯ Pansexual ◯ Asexual

4. Eres una persona en situación de discapacidad:

◯ Sí ◯ No

5. En caso de responder sí a la pregunta anterior, por favor indica el tipo de discapacidad:

◯ Motriz ◯ Visual ◯ Auditiva ◯ Intelectual

◯ Psicosocial ◯ No aplica

6. Máximo nivel de estudios (selección única NO no múltiple)

◯ Sin estudios

◯ Primaria Completa ◯ Primaria Incompleta

◯ Secundaria Completa ◯ Secundaria Incompleta

◯ Bachillerato Completa ◯ Bachillerato Incompleta

◯ Profesional Completa ◯ Profesional Incompleta

◯ Posgrado

7. Estado de residencia

(lista desplegable con el nombre de cada estado)

8. Principal ocupación laboral _____

b) Usos y prácticas en redes sociales

9. ¿Qué redes sociales usas? _____

10. ¿Con qué frecuencia utilizas las redes sociales?

◯ A diario ◯ Semanalmente

◯ Cuando tengo tiempo libre ◯ No uso redes sociales

11. ¿Cuántas horas al día utilizas las redes sociales?

◯ Menos de 5h ◯ De 5 a 10 h ◯ Siempre estoy conectado

12. Enumera del 1 al 10 para qué utilizas las redes sociales *(el 1 es la principal actividad y 10 es la actividad que menos realizas en redes sociales)*

___ Ver contenido informativo/periodístico

___ Colgar contenido e información personal

___ Colgar contenido e información de la organización/ colectiva

___ Colgar contenido informativo/periodístico/investigación

___ Opinar sobre temas varios

___ Opinar sobre tu causa

___ Prefiero no decir

___ Para buscar información

___ Para para asociarse o reunirse

___ Otra _____

13. Habitualmente, ¿generas opiniones a través de las redes sociales?

◯ Sí ◯ No

14. Indica cuánto tiempo ha transcurrido desde tu última publicación/opinión en redes sociales.

◯ Unas horas ◯ Unos días/semanas ◯ Hace bastante/mucho tiempo

15. De acuerdo con tu situación de discapacidad ¿consideras que la accesibilidad de las redes sociales que utilizas es?:

◯ Insuficiente ◯ Suficiente ◯ No aplica

16. ¿No usas o has dejado de usar redes sociales porque su accesibilidad y usabilidad es insuficiente dada tu situación de discapacidad?

◯ Sí ◯ No ◯ No aplica

c) Experiencias sobre libertad de expresión (en general)

17. ¿Te ha pasado que quieres hacer una publicación o compartir una opinión en tus redes sociales, pero finalmente no lo publicas por miedo a las reacciones en la red?

◯ Sí, me ha pasado ◯ No me ha pasado

18. ¿Has conocido alguna persona que tiene la intención de emitir un mensaje en las redes sociales, pero finalmente no lo publica por miedo a las reacciones en la red?

◯ Sí, conozco a gente ◯ No conozco gente

19. Cuando alguien expresa opiniones diametralmente opuestas a las tuyas en las redes sociales, ¿Qué sueles hacer?

◯ Interactuar con emoticonos ◯ Ignorarlas y no contestar

◯ Posteo y rebato la opinión ◯ Enojarme y no contestar

20. ¿Consideras que en general las personas usuarias respetan las opiniones en las redes sociales y se produce un debate educado y respetuoso?

◯ Sí ◯ No

21. ¿Por qué lo consideras así? *(pregunta abierta)*

22. ¿Quién debe ser responsable de las opiniones vertidas en la red?

◯ La Plataforma ◯ La persona usuaria ◯ Ambos

23. ¿Conoces o has leído las normas comunitarias de las redes sociales?

◯ Las he leído ◯ He escuchado hablar de ellas ◯ No las he leído

d) Experiencias sobre libertad de expresión II (moderación de contenido)

24. Compártenos qué es lo conoces sobre las normas comunitarias

(pregunta abierta)

25. ¿Has pasado por alguna de estas experiencias? *(Marca todas las que consideres)*

☐ Me bloqueó la plataforma por contenido inapropiado

☐ Me bloqueó mi cuenta personal por contenido que no cumplía con las normas comunitarias

☐ Me bloqueó la plataforma, pero no me explicó el motivo

☐ Eliminó contenido por no cumplir las normas comunitarias

☐ Eliminó contenido por no cumplir las normas y cuando pedí revisión no hubo respuesta o fue en negativo

☐ Eliminó contenido, pero no me explicó el motivo

☐ Bajaron mi cuenta personal

☐ Bajaron la cuenta de la organización/colectivo de la que formo parte

☐ Discursos estigmatizantes, o de incitación al odio

☐ Mensajes privados de amenazas

☐ Extorsión

☐ Suplantación de mi cuenta personal

26. En caso de que sí hayas vivido alguna de las experiencias anteriores ¿Qué tipo de medidas tomaste?

27. En caso de pasar por alguna de las experiencias anteriores, ¿Cuentas con una red de apoyo?

◯ Sí ◯ No

28. En caso de pasar por alguna de las experiencias anteriores, ¿cuentas con el apoyo de organizaciones, colectivos o personas que puedan asesorarte, acompañarte o apoyarte en caso de ser necesario?

◯ Sí ◯ No

29. En caso de responder Sí a la pregunta anterior, nombra a las organizaciones, colectivos o personas en quien te apoyarías.

30. ¿A quién no recurrirías en caso de pasar por una experiencia como las planteadas en las preguntas anteriores?

☐ Organización de la Sociedad Civil

☐ Policía Cibernética

☐ Fiscalía

☐ Institución pública

- ☐ Centro Laboral
- ☐ Centro Educativo
- ☐ Familia
- ☐ Amistades
- ☐ Pareja
- ☐ Otro _____

31. En relación con tu respuesta en la pregunta anterior ¿Por qué no recurrirías a esa persona, institución, organización o dependencia?

32. ¿Cuáles son las principales razones por las que te generan desconfianza las instituciones mencionadas en la pregunta anterior?

33. ¿Consideras que tu género influyó en cómo es que se dio esa experiencia?

◯ Sí ◯ No

34. En relación con tu respuesta en la pregunta anterior ¿Por qué lo consideras así?

35. Consideras que tú seguridad digital es:

◯ Buena ◯ Insuficiente ◯ Suficiente

36. ¿Crees que el contenido que se muestra en tus redes sociales es un reflejo de tus gustos y afinidades, investigaciones o información que compartes?

◯ Sí ◯ No

37. ¿Has identificado noticias falsas en las redes sociales?

◯ Sí, frecuentemente ◯ Muy pocas

◯ No las he visto ◯ No podría decir si es una noticia falsa

38. ¿Qué haces si identificas una noticia falsa?

◯ Reporto ◯ Ignoro ◯ La leo para saber de qué trata

39. ¿La colectiva u organización de la que eres parte, cuenta con el apoyo de alguna organización para consultar o asesorar sobre derechos a la libertad de expresión y acceso a la información en plataformas digitales?

◯ Sí ◯ No

40. En caso de responder Sí a la pregunta anterior, nombra a las organizaciones, colectivos o personas en quien te apoyarías.

Anexo 3. Guía de entrevista

Guía General para entrevistas a especialistas

General	Tener una aproximación sobre los principales problemas que enfrentan grupos vulnerables en el ejercicio de sus derechos a la libertad de expresión y acceso a la información en las plataformas digitales.
Específicas	Atiende a la experiencia y trayectoria particular del entrevistado. Busca profundizar en aquellos tópicos que por su perfil resulta especialmente valioso conocer sus respuestas.
De control	Pregunta que verifica las respuestas a las otras y su consistencia interna.

Puntos Generales

01 Vinculación del eje temático con las plataformas digitales de redes sociales
02 Problemáticas:
- **a** Acceso a la información (infraestructura y conectividad)
- **b** Libertad de expresión (censura y moderación de contenido)

03 Casos concretos y experiencias
04 Acciones para el ejercicio de los derechos digitales:
- **a** Inclusión
- **b** Visibilización

05 Recomendaciones

Preguntas Específicas

Migrantes

01 ¿Qué relación existe entre las plataformas de redes sociales y los procesos migratorios?
02 ¿Cuáles son los principales problemas que enfrentan los migrantes en relación con la infraestructura de Internet (acceso a la información)?

03 ¿Cuáles son los principales problemas que enfrentan los migrantes en relación con la libertad de expresión en las plataformas de redes sociales?
04 ¿Qué afectación tiene la desinformación (*fakenews*) y los discursos estigmatizantes y de odio en los grupos migrantes?
05 ¿Puede la moderación de contenido automatizada afectar la interacción de la migración en México?
06 ¿Cómo?
07 ¿Cómo las redes sociales han aportado a visibilizar la problemática migratoria?

Personas con discapacidad

01 ¿Qué relación existe entre las plataformas digitales y las personas con discapacidad?
02 ¿Cuáles son los principales problemas que enfrentan las personas con discapacidad en relación con la infraestructura de Internet (acceso a la información)?
03 ¿Cuáles son los principales problemas que enfrentan las personas con discapacidad en relación con la libertad de expresión en las plataformas de redes sociales?
04 ¿Qué afectación tiene la desinformación (*fakenews*) y los discursos estigmatizantes y de odio en los grupos migrantes?
05 ¿Puede la moderación de contenido automatizada afectar a las personas con discapacidad en México?
06 ¿Cómo?
07 ¿Pueden servir las plataformas de redes sociales como una herramienta para la visualización de las personas con discapacidad?

Comunidades Indígenas

01 ¿Qué relación existe entre las plataformas digitales y los pueblos indígenas?
02 ¿Cuáles son los principales problemas que enfrentan las comunidades indígenas en relación con la infraestructura de Internet (acceso a la información)?
03 ¿Cuáles son los principales problemas que enfrentan las comunidades indígenas en relación con la libertad de expresión en las plataformas de redes sociales?
04 ¿Qué afectación tiene la desinformación (*fakenews*) y los discursos estigmatizantes y de odio a las comunidades indígenas?
05 ¿Puede la moderación de contenido automatizada afectar a las comunidades indígenas?
06 ¿Cómo?

07 ¿Pueden servir las plataformas de redes sociales como una herramienta para la visualización de las personas con discapacidad?

LGBTQ+

01 ¿Qué relación existe entre las plataformas digitales y la comunidad LGBTG+?
02 ¿Cuáles son los principales problemas que enfrenta la comunidad LGBTG+ en relación con la infraestructura de Internet (acceso a la información)?
03 ¿Cuáles son los principales problemas que enfrenta la comunidad LGBTG+ en relación con la libertad de expresión en las plataformas de redes sociales?
04 ¿Puede la remoción, bloqueo, baja de contenido o cuentas afectar derechos de la comunidad LGBTG+?
05 ¿Qué afectación tiene la desinformación (*fakenews*) y los discursos estigmatizantes y de odio a la comunidad LGBTG+?
06 ¿Puede la moderación de contenido automatizada afectar a la comunidad LGBTG+ en México?
07 ¿Cómo?
08 ¿Pueden servir las plataformas de redes sociales como una herramienta para la visualización de la comunidad LGBTG+?

Mujeres

01 ¿Qué relación existe entre las plataformas digitales y las mujeres?
02 ¿Cuáles son los principales problemas que enfrentan las mujeres en relación con la infraestructura de Internet (acceso a la información)?
03 ¿Cuáles son los principales problemas que enfrentan las mujeres en relación con la libertad de expresión en las plataformas de redes sociales?
04 ¿Puede la remoción, bloqueo, baja de contenido o cuentas afectar derechos de las mujeres?
05 ¿Qué afectaciones producen la desinformación (*fakenews*) y los discursos estigmatizantes y de odio a las mujeres?
06 ¿Puede la moderación de contenido automatizada afectar a las mujeres en México?
07 ¿Cómo?
08 ¿Pueden servir las plataformas de redes sociales como una herramienta para la visualización de las mujeres?