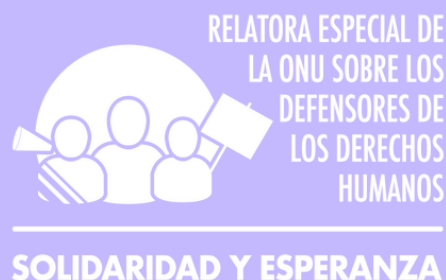


DOCUMENTO DE
ALCANCE

**De la visibilidad a la
vulnerabilidad:
cuando las redes
sociales fallan a las
personas defensoras
de los derechos
humanos**

Marzo de 2026



ÍNDICE

Introducción	03
Metodología	04
Derecho y normas internacionales	05
Fallos de moderación de contenidos y visibilidad comprometida	06
Cumplimiento con peticiones represivas de gobiernos	11
Ataques digitales y la respuesta inadecuada de las plataformas	13
Tratamiento desigual de usuarios y contenidos	17
Dificultades al interactuar con las plataformas de redes sociales	18
Conclusión	23
Recomendaciones	25

INTRODUCCIÓN

Desde su creación, las plataformas de redes sociales han pasado a ser herramientas indispensables para las personas que defienden los derechos humanos. La gente recurre a ellas para obtener información, despertar conciencia sobre éxitos y retos, compartir informes y documentación, construir y organizar comunidades y redes, y amplificar voces marginadas.

Sin embargo, en el transcurso de su labor, y en consultas con personas que defienden los derechos humanos, la Relatora Especial de la ONU sobre la situación de las personas defensoras de derechos humanos (“la Relatora Especial”) ha constatado que estas mismas plataformas obstaculizan de manera significativa el activismo en defensa de los derechos humanos. Mediante prácticas de moderación que en el mejor de los casos son defectuosas, supresión algorítmica, respuestas inadecuadas al acoso en línea y el cumplimiento de las demandas represivas de Estados, las plataformas de redes sociales contribuyen, ya sea intencionadamente o no, al silenciamiento, *deplatforming* (veto y bloqueo deliberado de acceso a plataformas) y la puesta en peligro de las personas defensoras de los derechos humanos (“DDH”). [1]

Dadas estas preocupaciones, la Relatora Especial publica este documento de alcance para examinar el impacto de las redes sociales sobre el derecho a defender los derechos humanos. Sus hallazgos subrayan la necesidad urgente de que las empresas detrás de las plataformas de redes sociales cumplan con sus responsabilidades en materia de derechos humanos, y de que los Estados respeten sus obligaciones relativas a los derechos humanos y se aseguren de que los espacios digitales continúen siendo entornos seguros y favorables para aquellas personas que defienden los derechos humanos.

[1] La Relatora Especial define una persona defensora de los derechos humanos como una persona que, individualmente o con otras personas, actúa de manera pacífica para promover o proteger los derechos humanos, según la Declaración de las Naciones Unidas sobre la situación de las personas defensoras de los derechos humanos.

METODOLOGÍA

Este documento de alcance se basa en investigaciones y consultas [2] realizadas por la oficina de la Relatora Especial de la ONU sobre la situación de las personas defensoras de los derechos humanos con organizaciones de la sociedad civil que trabajan con personas que defienden y promueven los derechos humanos en todas las regiones.

El alcance del documento tiene limitaciones de tiempo, recursos y acceso, y no pretende ser exhaustivo. Más bien, refleja las experiencias de aquellas organizaciones que pudieron contribuir al proceso de consulta. Debe considerarse como un mapeo indicativo de patrones y riesgos clave, y no como una evaluación comprehensiva de todos los impactos relativos a las plataformas sobre el derecho a defender los derechos humanos.

La Relatora Especial de la ONU sobre la situación de las personas defensoras de los derechos humanos compartió sus preocupaciones acerca de las políticas de [Meta Inc.](#) y [X Corp.](#) y su impacto sobre el derecho a defender los derechos humanos en dos comunicados oficiales enviados a las empresas el 30 de diciembre de 2025, informando que los comunicados y sus respuestas a ellos se incluirían en este documento de alcance. A día de la publicación de este informe, no se había recibido ninguna respuesta.

[2] Las consultas se realizaron bajo las reglas de Chatham House.

1. DERECHO Y NORMAS INTERNACIONALES

Adoptada por consenso en la Asamblea General de la ONU de diciembre de 1998, la Declaración de la ONU sobre los defensores de los derechos humanos [3] (“la Declaración DDH”) declara en el artículo 1 que “toda persona tiene derecho, individual o colectivamente, a promover y procurar la protección y realización de los derechos humanos y las libertades fundamentales en los planos nacional e internacional.” Esto se refuerza en el artículo 12.1, que declara que “toda persona tiene derecho, individual o colectivamente, a participar en actividades pacíficas contra las violaciones de los derechos humanos y las libertades fundamentales”, mientras que el artículo 6 sostiene que toda persona tiene derecho a publicar, impartir o difundir libremente a terceros opiniones, informaciones y conocimientos relativos a todos los derechos humanos y las libertades fundamentales.

El artículo 6 de la Declaración DDH desarrolla el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos, que protege el derecho a la libertad de opinión y la libertad de expresión, definido como la “libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección”. Según el artículo 19(3), la libertad de expresión puede estar sujeta a restricciones, que deben ser claras y precisas (“expresamente fijadas por la ley”) y deben ser necesarias para asegurar “el respeto a los derechos o a la reputación de los demás”, “la protección de la seguridad nacional, el orden público”, o “la salud o la moral públicas”.

Los estados son los principales titulares de deberes bajo el derecho internacional de los derechos humanos, y deben proteger frente a violaciones de derechos humanos dentro de su territorio o jurisdicción por parte de terceros, incluyendo empresas. Sin embargo, bajo los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos, las empresas tienen la responsabilidad de: a) evitar que sus propias actividades provoquen o contribuyan a provocar consecuencias negativas sobre los derechos humanos; y (b) prevenir o mitigar tales consecuencias negativas directamente relacionadas con operaciones, productos o servicios prestados por sus relaciones comerciales. Para cumplir con su responsabilidad de respetar los derechos humanos, las empresas deben tener a) un compromiso político de asumir su responsabilidad de respetar los derechos humanos; b) un proceso de diligencia debida en materia de derechos humanos para

[3] A/RES/53/144, disponible en múltiples idiomas en: <https://www.ohchr.org/en/special-procedures/sr-human-rights-defenders/declaration-human-rights-defenders-different-languages>

identificar, prevenir, mitigar y rendir cuentas de cómo abordan su impacto sobre los derechos humanos; y c) unos procesos que permitan reparar todas las consecuencias negativas sobre los derechos humanos que hayan provocado o contribuido a provocar. La responsabilidad de respetar los derechos humanos constituye una norma mundial de conducta aplicable a todas las empresas y existe independientemente de la capacidad o voluntad de los Estados de cumplir con sus propias obligaciones en materia de derechos humanos.

2. FALLOS DE MODERACIÓN DE CONTENIDOS Y VISIBILIDAD COMPROMETIDA

Las personas defensoras de los derechos humanos en todas las regiones informan de varios fallos de moderación de contenido en plataformas de redes sociales, con impactos graves sobre el derecho a defender los derechos humanos. Estos incluyen eliminación de contenidos legítimos, bloqueo en la sombra o *shadowbanning*, retiradas arbitrarias, aplicación desigual de las directrices comunitarias, y una falta generalizada de transparencia en los procesos internos.

2.1 ELIMINACIÓN DE CONTENIDOS LEGÍTIMOS SOBRE DERECHOS HUMANOS

La información recibida para la preparación de este informe incluye incidentes informados por varias organizaciones, en los que se habían eliminado contenidos legítimos sobre derechos humanos por supuestamente violar las políticas de la plataforma.

Una organización informó que las publicaciones relacionadas con los derechos humanos en **Palestina** habían sido retiradas por plataformas pertenecientes a Meta, por haber sido categorizadas bajo "Individuos y organizaciones peligrosas" o "Contenidos violentos," a menudo sin una clara explicación ni opción a recurrir. Estas publicaciones incluían documentación de matanzas extrajudiciales, detenciones arbitrarias y protestas pacíficas, y las retiradas afectaron especialmente a personas defensoras de los derechos humanos y periodistas, algunas de las cuales no recibieron ninguna notificación ni explicación de la censura. Las personas DDH también informaron de inconsistencias: por ejemplo, se retiraron publicaciones que criticaban contenidos de odio o extremistas, mientras que el contenido de odio o extremista original que criticaban dichas publicaciones permaneció en

en línea [4]. Una organización de derechos digitales también documentó de manera sistemática la censura y retirada de contenidos relativos a los derechos humanos en Palestina por parte de plataformas de redes sociales, incluyendo Facebook, X e Instagram [5]. Asimismo, otra organización informó de retiradas de contenidos que documentaban casos de víctimas de la guerra en **Siria**, incluyendo publicaciones que podían haber contribuido a los esfuerzos internacionales de rendición de cuentas [6]. Las razones que generalmente proporcionan las plataformas para justificar la retirada de tales contenidos incluyen ensalzar a individuos u organizaciones peligrosas, contenidos violentos o gráficos. Las fotografías de víctimas, especialmente en el caso de menores, se han marcado como explotación infantil.

También se han visto problemas similares en otras regiones y en otras plataformas. En TikTok, contenidos relativos a abusos de derechos humanos en **Xinjiang** y **Tíbet** han sido repetidamente marcados, retirados o restringidos en los últimos años por supuestamente infringir las directrices comunitarias [7][8]. Una organización de derechos humanos bielorrusa informó que sus publicaciones sobre víctimas de la represión política en **Bielorrusia** a veces eran marcadas como contenido inapropiado o político en Facebook e Instagram, y que sus apelaciones obtuvieron diferentes resultados, con pocas explicaciones [9]. En **Camboya**, las cuentas de Facebook de periodistas han sido suspendidas tras haber publicado contenidos sobre disputas de tierras, corrupción y deforestación, y durante las elecciones nacionales de 2023 varias cuentas de medios de comunicación fueron bloqueadas, gravemente impidiendo el acceso a información independiente [10]. Una organización regional de derechos humanos informa que una de sus campañas pagadas sobre ASEAN y los derechos humanos fue marcada como contenido inapropiado o delicado y fue retirada de las plataformas de las redes sociales en **Filipinas** y **Camboya**, sin una justificación clara y a pesar de estar activa sin ninguna incidencia en otros países asiáticos [11].

[4] Comunicación (A) recibida por la Relatora Especial

[5] Comunicación (B) recibida por la Relatora Especial

[6] Comunicación (C) recibida por la Relatora Especial

[7] *The Guardian*, “Revealed: how TikTok censors videos that do not please Beijing” (*Revelamos: cómo TikTok censura videos que no agradan a Pekín*), septiembre de 2019 <https://www.theguardian.com/technology/2019/sep/25/revealed-how-tiktok-censors-videos-that-do-not-please-beijing>

[8] *Radio Free Asia*, “TikTok deletes videos related to Uyghur human rights violations” (*TikTok elimina videos relacionados con violaciones de los derechos humanos de los uygur*), noviembre de 2024 <https://www.rfa.org/english/uyghur/2024/11/05/uyghur-tiktok-censors-abroad/>

[9] Comunicación (D) recibida por la Relatora Especial

[10] Comunicación (E) recibida por la Relatora Especial

[11] Comunicación (F) recibida por la Relatora Especial

2.2 SHADOWBANNING Y LA SUPRESIÓN DE VISIBILIDAD

El bloqueo en la sombra o *shadowbanning*, cuando las publicaciones de los usuarios se vuelven invisibles para otros usuarios sin previo aviso, es un problema ampliamente denunciado por las personas defensoras de los derechos humanos.

Varias organizaciones han constatado que las cuentas de personas defensoras de los derechos humanos palestinos y contenidos relativos a los derechos humanos en **Palestina** habían sido sometidas a *shadowbanning*, especialmente en las plataformas de Meta, Facebook e Instagram, y que estas prácticas han empeorado desde el inicio del genocidio que está perpetrando Israel [12], aunque también se empleaban anteriormente [13]. Esto típicamente significa que la información compartida por estas personas DDH, sea para documentar violaciones de derechos humanos sobre el terreno, crear conciencia internacional o recaudar fondos para apoyar a víctimas, alcanzará a menos personas, y por tanto su labor será menos eficaz. Los usuarios afectados no reciben ni notificaciones ni justificaciones. Tales casos pueden ser difíciles de comprender y verificar, ya que Meta proporciona poca transparencia sobre la aplicación de sus algoritmos [14][15].

Organizaciones que trabajan en la libertad de expresión han encontrado que el *shadowbanning* ha afectado de manera desproporcionada a mujeres feministas defensoras de los derechos humanos en el **Líbano**, quienes informan que se han retirado sus publicaciones, especialmente en idioma árabe, mientras que siguen estando disponibles contenidos de odio hacia las mujeres [16][17]. Las defensoras de derechos humanos que defienden los derechos sexuales y reproductivos también se ven afectadas [18].

En un ejemplo documentado por una de las organizaciones que presentaron comunicaciones, un colectivo **chileno** de periodistas migrantes informó que las publicaciones sobre un evento antirracista en Instagram no estaban apareciendo en los timelines. A pesar de peticiones repetidas de ayuda, no se proporcionó ninguna explicación clara de por qué estaban escondidas las publicaciones, y la visibilidad

[12] Amnistía Internacional, “Después del alto al fuego: Continúa el genocidio de Israel en la Franja de Gaza ocupada”, 27 de noviembre de 2025 <https://www.amnesty.org/es/documents/mde15/0527/2025/es/>

[13] SMEX, “Israel’s Fierce Attack on Palestinian Narrative and Content” (*El feroz ataque de Israel a la narrativa y los contenidos palestinos*), junio de 2022 <https://smex.org/israels-fierce-attack-on-palestinian-narrative-and-content/>

[14] Comunicación (A) recibida por la Relatora Especial

[15] Comunicación (B) recibida por la Relatora Especial

[16] Comunicación (B) recibida por la Relatora Especial

[17] SMEX, “Research Report: Confronting Structural Silencing: Challenges and Resistance Among Digital Feminist Activists in Lebanon” (*Informe de investigación: Enfrentarse al silenciamiento estructural: Retos y resistencia entre activistas feministas digitales en Líbano*), abril de 2025 <https://smex.org/research-report-confronting-structural-silencing-challenges-and-resistance-among-digital-feminist-activists-in-lebanon/>

[18] Comunicación (C) recibida por la Relatora Especial

únicamente se restableció después de largos intercambios con las plataformas. La falta de transparencia dejó al colectivo con incertidumbre acerca de las razones detrás de las restricciones y lo disuadió de crear más publicaciones [19].

El *shadowbanning* también se ha usado en detrimento de personas implicadas en la defensa de los derechos LGBTIQ+. En 2024, cambios en la política [20] de Meta limitaron algorítmicamente el alcance de “contenidos políticos” en Instagram y Threads, sin aclarar exactamente qué significaría esto y cómo se aplicaría. En términos reales, este cambio redujo de manera drástica el alcance de publicaciones creíbles e informativas de varias cuentas de Instagram en defensa de los derechos LGBTIQ+ [21], incluyendo las cuentas de las organizaciones **estadounidenses** Human Rights Campaign y GLAAD. Este desarrollo agravó el largo recorrido por parte de Meta de realizar *shadowbanning* a contenidos LGBTIQ+, a menudo marcándolos como sexualmente explícitos [22].

2.3 FALLOS DEL ALGORITMO Y DE LA MODERACIÓN AUTOMÁTICA

Muchos de los fallos de moderación mencionados anteriormente parecen derivar del uso de sistemas de moderación automáticos o impulsados por IA. Según una organización no gubernamental, tales tecnologías a menudo carecen de la complejidad necesaria para poder distinguir de manera fiable entre contenidos legales e ilegales, lo cual lleva a la eliminación de enormes cantidades de expresiones legítimas, incluidas por parte de personas defensoras de derechos humanos [23].

Otra organización citó un ejemplo en el que los filtros automatizados habían clasificado erróneamente publicaciones legítimas que documentaban la represión durante las elecciones de 2024 en **Venezuela** como “incitación a la violencia”, lo que les llevó a ser bloqueadas [24]. También informaron que se suspendió permanentemente la cuenta de un colectivo de derechos reproductivos en **Brasil**, por compartir información sobre Misoprostol, un medicamento para úlceras que había sido usado desde hace mucho por

[19] Comunicación (A) recibida por la Relatora Especial

[20] *Instagram*, “Actualización sobre cómo abordamos el contenido político en Instagram y Threads”, febrero de 2024 <https://about.instagram.com/es-la/blog/announcements/continuing-our-approach-to-political-content-on-instagram-and-threads>

[21] *Accountable Tech*, “Meta’s Political Content Limit Causes Steep Drop in Reach for Accounts”, julio de 2024 <https://accountabletech.org/research/metas-political-content-limit-causes-steep-drop-in-reach-for-accounts/>

[22] *Human Rights Campaign*, “Meta’s New Policies: How They Endanger LGBTQ+ Communities and Our Tips for Staying Safe Online” (*Las nuevas políticas de Meta: cómo ponen en peligro a las comunidades LGBTQ+ y consejos de seguridad en línea*), enero de 2025 <https://www.hrc.org/news/metas-new-policies-how-they-endanger-lgbtq-communities-and-our-tips-for-staying-safe-online>

[23] Comunicación (G) recibida por la Relatora Especial

[24] Comunicación (A) recibida por la Relatora Especial

mujeres brasileñas para inducir abortos [25].

2.4 SUSPENSIONES DE CUENTAS Y DENUNCIAS MASIVAS

Las personas y organizaciones de la sociedad civil que defienden los derechos humanos también se enfrentan a suspensiones de cuentas y restricciones desencadenadas por campañas de denuncias masivas coordinadas, realizadas por *trolls*, agentes vinculados a estados, grupos de derecha y grupos antiderechos.

En **Centroamérica**, una organización regional señaló el uso de *bots* y campañas de abuso coordinadas para generar denuncias falsas, lo que llevó a retiradas de publicaciones y cierres de cuentas para personas que defienden los derechos humanos. Estos incidentes a menudo ocurren sin previo aviso ni explicación, y los procesos de apelación son opacos e inaccesibles, especialmente para usuarios que no tienen acceso directo a los canales de Socios de Confianza de las plataformas [26]. En **Bangladés**, personas defensoras de los derechos humanos feministas, *queers* y laicas experimentaron problemas parecidos después de que campañas de denuncias masivas coordinadas resultaran en suspensiones automáticas de sus cuentas en Facebook e Instagram, sin ningún recurso significativo [27]. En **Irán**, una organización informó que las cuentas de Instagram de mujeres activistas de derechos de la mujer iraníes fueron cerradas o bloqueadas en la sombra (*shadowbanning*) tras denuncias masivas atribuidas a las autoridades iraníes [28]. En **Bélgica**, una cuenta de Instagram dedicada a la educación antirracista y administrada por una joven mujer defensora de los derechos humanos, fue desactivada siete veces entre 2023 y 2025 debido a denuncias masivas de grupos antiderechos, lo cual le llevó en última instancia a poner la cuenta en modo privado y dejar de subir nuevas publicaciones [29]. En **Malasia**, una organización relata que las cuentas de Facebook e Instagram de una organización de derechos de la mujer y dos de sus miembros de personal fueron suspendidas sin explicación, tras una ola de acoso en línea. Aunque las cuentas fueron posteriormente reactivadas, el incidente afectó su labor de comunicación, y la organización tuvo que realizar una revisión interna de su estrategia digital [30].

[25] Comunicación (A) recibida por la Relatora Especial

[26] Comunicación (H) recibida por la Relatora Especial

[27] Comunicación (A) recibida por la Relatora Especial

[28] Comunicación (B) recibida por la Relatora Especial

[29] <https://www.instagram.com/p/DHvS29aomok/>

[30] Comunicación (G) recibida por la Relatora Especial

3. CUMPLIMIENTO CON PETICIONES REPRESIVAS DE GOBIERNOS

Al igual que las personas que usan las plataformas de las redes sociales para la promoción y protección de los derechos humanos se enfrentan a interferencia de gobiernos que buscan silenciar el disenso y controlar las narrativas [31], algunas empresas frecuentemente acceden a peticiones de censura por parte de gobiernos, que son contrarias al derecho de los derechos humanos o facilitan la vigilancia por parte de los estados.

3.1 CUMPLIMIENTO CON PETICIONES DE CENSURA

En información recibida por la Relatora Especial en preparación para este informe, varias organizaciones destacaron instancias en las que los Gobiernos ejercieron presión sobre las plataformas de redes sociales para alinear sus prácticas de moderación de contenidos con puntos de vista políticos específicos o para censurar directamente a personas defensoras de los derechos humanos. Tal presión puede tomar la forma de amenazas de *throttling* o estrangulamiento (limitar el ancho de banda disponible en la plataforma), prohibiciones publicitarias o sanciones legales como multas.

En **Türkiye**, donde las plataformas de redes sociales se han convertido en la fuente principal de noticias y una herramienta vital para el periodismo independiente, en los últimos años el Gobierno ha arreciado su control sobre las plataformas de redes sociales mediante una serie de enmiendas a la Ley de Internet. Una organización informó de tasas muy altas de cumplimiento con peticiones de eliminación de contenido por parte del gobierno turco en las principales plataformas de redes sociales: un 95% en el caso de TikTok, un 85% en el caso de X y un 79% en el caso de Instagram en la segunda mitad de 2024) [32]. En torno a la detención del líder de la oposición Ekrem İmamoğlu en marzo de 2025, el gobierno ordenó el bloqueo de cientos de cuentas, entre ellas cuentas de personas defensoras de derechos humanos, periodistas, grupos estudiantiles, organizaciones de la sociedad civil y grupos de derechos de la mujer. Si bien X cumplió con muchas de estas órdenes de bloqueo debido a la amenaza de estrangulamiento, Meta declaró que no tomó ninguna acción [33] y según se informa tuvo que hacer frente a una

[31] Véase [A/HRC/38/35](#), párrafo 13, [A/80/341](#) párrafos 12 y 76.

[32] Comunicación (G) recibida por la Relatora Especial

[33] Meta, "Case Studies - Content alleged to violate Turkish Law" (*Estudios de caso: contenidos que supuestamente infringen la ley turca*), marzo de 2025 <https://transparency.meta.com/reports/content-restrictions/case-studies>

multa elevada [34]. La organización también informó que Meta [35], TikTok [36], YouTube [37] y X [38] fueron todos presionados para nombrar a representantes en el país, bajo la amenaza de crecientes repercusiones legales, a pesar de avisos de que estos darían al gobierno turco un mayor control sobre sus plataformas [39].

En **India**, *Bolta Hindustan*, una plataforma mediática independiente que se especializa en temas sociales y tiene 300.000 suscriptores, fue bloqueada en YouTube en abril de 2024 tras una notificación del Ministerio de Información y Radiodifusión. Inicialmente, el equipo legal de Google no proporcionó a la plataforma mediática ninguna razón por el bloqueo [40]. Defensores de la libertad de prensa denunciaron el bloqueo, y la cuenta fue reactivada posteriormente por YouTube [41]. Por su parte, Twitter (ahora X) ha sido acusada de censurar voces de Cachemira a petición del gobierno indio. En 2019, el Comité para la Protección de los Periodistas informó que Twitter impedía a usuarios en India acceder a miles de tuits donde se compartía información acerca de la situación en **Jammu y Cachemira** durante bloqueos mediáticos y apagones de Internet, tras avisos legales enviados por el gobierno indio [42].

En **Pakistán**, retransmisiones en vivo de las protestas baluchis en Facebook y YouTube se vieron afectadas por una velocidad menor de Internet, o simplemente se eliminaron sin explicación [43].

[34] Associated Press, "Meta is fined by Turkey after refusing to restrict content on Facebook and Instagram" (*Meta es multada por Turquía tras negarse a restringir contenidos en Facebook e Instagram*), abril de 2025

<https://apnews.com/article/meta-facebook-instagram-turkey-erdogan-59f16571e884fd6f0f9597a72771131f>

[35] Facebook, "An Update on Facebook in Turkey" (*Actualización sobre Facebook en Turquía*), enero de 2021

<https://about.fb.com/news/2021/01/an-update-on-facebook-in-turkey/>

[36] TikTok, "An update on TikTok in Turkey" (*Actualización sobre TikTok en Turquía*), enero de 2021

<https://newsroom.tiktok.com/an-update-on-tiktok-in-turkey?lang=en>

[37] YouTube, "An update on YouTube in Turkey" (*Actualización sobre YouTube en Turquía*), diciembre de 2020 <https://blog.youtube/news-and-events/update-youtube-turkey/>

[38] *Balkan Insight*, "X Appoints Representative in Turkey, Bowing to Govt Pressure" (*X cede a la presión del Gobierno y nombra a un representante en Turquía*), mayo de 2024 <https://balkaninsight.com/2024/05/23/x-appoints-representative-in-turkey-bowing-to-govt-pressure/>

[39] Comunicación (G) recibida por la Relatora Especial

[40] Comunicación (F) recibida por la Relatora Especial

[41] *Medianama*, "YouTube Reinstates Blocked and Demonetised Channels" (*YouTube reactiva canales bloqueados y desmonetizados*), mayo de 2024 <https://www.medianama.com/2024/05/223-youtube-reinstates-blocked-demonetised-channels/>

[42] *Committee to Protect Journalists*, "India uses opaque legal process to suppress Kashmiri journalism, commentary on Twitter" (*India utiliza procesos legales opacos para suprimir el periodismo y los comentarios cachemires en Twitter*), octubre de 2018 <https://cpj.org/2019/10/india-opaque-legal-process-suppress-kashmir-twitter/>

[43] Comunicación (F) recibida por la Relatora Especial

3-2 ENTORNO FAVORABLE PARA LA VIGILANCIA GUBERNAMENTAL

Otra preocupación es la falta de transparencia y control sobre datos personales compartidos por personas defensoras de los derechos humanos en sus páginas de redes sociales. En casos documentados por una de las organizaciones que presentó comunicaciones para este documento, los datos de usuario pertenecientes a personas defensoras de los derechos humanos habían sido supuestamente proporcionados a las autoridades estatales sin su conocimiento ni consentimiento. Por ejemplo, en **Guatemala**, Meta compartió los datos de usuario de una persona defensora de los derechos humanos, no solo relativos a la cuenta requerida por las autoridades, sino de todas sus cuentas de redes sociales. La persona defensora nunca fue notificada oficialmente al respecto, y solo supo de ello cuando una fuente de confianza que tenía conocimiento de los hechos le informó [44]. Esto plantea graves preocupaciones, especialmente en contextos en los que es conocido que las voces de disenso son perseguidas por los gobiernos. Los procesos mediante los cuales las plataformas evalúan y responden a las peticiones de las autoridades nacionales siguen siendo opacos. Como resultado, las personas que defienden los derechos humanos podrían estar sujetas a vigilancia, intimidación o represalias sin ser conscientes de que se han divulgado o se ha accedido a sus datos privados, y sin acceso a recursos efectivos.

4. ATAQUES DIGITALES Y LA RESPUESTA INADECUADA DE LAS PLATAFORMAS

Las plataformas de las redes sociales han reproducido los patrones interseccionales del acoso y el abuso a los que ya se enfrentan muchas personas defensoras de los derechos humanos en su trabajo, en una amplia gama de formas digitales. El abuso en línea se ha convertido tanto más prevalente como más coordinado, y las personas defensoras son blancos constantes de acoso, difamación y campañas de desinformación que se propagan en las plataformas de redes sociales y a menudo incluyen el uso de *bots* automatizados [45].

[44] Comunicación (A) recibida por la Relatora Especial

[45] Resolución 58/23 aprobada por el Consejo de Derechos Humanos el 4 de abril de 2025 <https://docs.un.org/es/A/HRC/RES/58/23> ; UN Women, “Tipping point: The chilling escalation of online violence against women in the public sphere” (*Punto crítico: la escalofriante escalada de violencia en línea contra mujeres en el ámbito público*), 2025, <https://www.unwomen.org/en/digital-library/publications/2025/12/tipping-point-the-chilling-escalation-of-violence-against-women-in-the-public-sphere-in-the-age-of-ai> ; Front Line Defenders, “Global Analysis 2024/2025”, https://www.frontlinedefenders.org/sites/default/files/1609_fld_ga24-5_output.pdf

Estas tácticas, a veces apoyadas por herramientas de inteligencia artificial, se están volviendo cada vez más prevalentes. Ante esta situación, las plataformas no proveen mecanismos adecuados para responder ni proteger a usuarios afectados por tales ataques.

Como ejemplo de la magnitud del problema, Global Witness destacó en un informe reciente [46] que el 92% de las personas defensoras de derechos humanos de tierras y medio ambiente que encuestaron habían sufrido abuso en línea por parte de una variedad de actores, incluidos individuos, bots y campañas coordinadas. A nivel global, las personas DDH recibieron más abuso en Facebook que en cualquier otra plataforma de redes sociales, seguida de X (Twitter) e Instagram, según el estudio [47].

4.1 ABUSOS EN LÍNEA BASADOS EN GÉNERO E IDENTIDAD

Las personas que ya sufren algún tipo de opresión suelen enfrentarse a riesgos mayores y abusos más duros cuando usan las plataformas de redes sociales para defender y promover los derechos humanos. En todas las regiones del mundo, las personas DDH, especialmente mujeres y defensores de derechos LGBTIQ+, se enfrentan a desinformación, campañas de desprestigio, difamación visual, *deepfakes*, extorsión, amenazas, acoso, *doxing* y suplantación de identidad, que a menudo son ignorados o abordados de manera inadecuada por las plataformas. El objetivo de estos ataques es deslegitimar, intimidar, desacreditar y ridiculizar a las personas defensoras y su labor.

Una organización en Centroamérica encontró que los ataques digitales suponían alrededor de una cuarta parte de todos los ataques contra mujeres DDH que documentó la organización entre 2020 y junio de 2025, un número que ha incrementado continuamente cada año. En **El Salvador**, los ataques digitales representaron más de la mitad de todos los ataques contra mujeres DDH. Casi un 90% de estos ataques se realizaron en redes sociales [48].

Otra organización informó que la defensora de derechos humanos siria Hiba Ezzideen Al-Hajji sufrió una campaña de difamación en Facebook cuyo objetivo era ella, su familia y la organización de la que forma parte, a raíz de una publicación contra el matrimonio forzado. Recibió numerosas amenazas de muerte a través de una avalancha de publicaciones en su propia cuenta, además de la página de su organización, con llamadas a la violencia física contra ella [49][50]. Unos años antes, ya había sido desprestigiada y amenazada a través

[46] Global Witness, “Plataformas tóxicas, planeta destruido”, 16 de julio de 2025

<https://globalwitness.org/es/campaigns/digital-threats/plataformas-toxicas-planeta-destruido/>

[47] Global Witness, *Ibid*

[48] Comunicación (H) recibida por la Relatora Especial

[49] Comunicación (A) recibida por la Relatora Especial

[50] Véase [AL SYR 3/2025](#)

de Telegram [51].

En **Bangladés**, una organización destacó que la transición política de 2024 marcó un punto de inflexión, cuando el acoso en línea evolucionó de incidentes aislados a campañas coordinadas, con una fuerte subida en *doxing*, difamación visual, discursos de odio basados en *memes*, denuncias masivas de cuentas de defensores de derechos humanos en Facebook e Instagram, y la circulación de desinformación focalizada. Estos ataques a menudo contenían abusos basados en género y religión, especialmente contra feministas, defensoras de derechos humanos laicos, y defensores de derechos humanos LGBTQI+ [52].

Varias organizaciones de derechos humanos informaron que el uso creciente de inteligencia artificial (IA) para generar *deepfakes* e imágenes manipuladas había aumentado la vulnerabilidad en línea de las personas defensoras de los derechos humanos, en particular las mujeres [53][54][55]. Este ha sido el caso en Venezuela, donde la defensora de los derechos humanos Martha Lía Grajales fue el blanco de una campaña de difamación basada en un vídeo generado por IA que promovía la falsa narrativa de que su labor de derechos humanos era una tapadera para un complot contra el gobierno. El vídeo fue compartido extensamente en redes sociales y WhatsApp, con la participación de milicias digitales dedicadas a realizar y amplificar tales ataques. Una vez se ha divulgado, es difícil eliminar o impugnar este tipo de contenido [56].

En todos los ejemplos mencionados anteriormente, las plataformas han sido lentas o ineficaces para responder a estos daños, según informes.

Múltiples comunicaciones recibidas destacaron el papel de las plataformas de redes sociales a la hora de alimentar y amplificar contenidos dañinos en base a sus algoritmos, que priorizan contenidos sensacionalistas, polarizantes y violentos y, por tanto, amplían el alcance de acosadores y abusadores [57][58][59]. Una organización advirtió que los algoritmos con fines de lucro que actualmente impulsan las principales plataformas de redes sociales contribuyen directamente a la violencia digital contra defensoras de los derechos humanos en **Mesoamérica**, y que son parte de un ciclo de violencia continuo que

[51] Véase [AL OTH 129/2023](#), al que los titulares de los mandatos de los Procedimientos Especiales no recibieron ninguna respuesta por parte de Telegram

[52] Comunicación (A) recibida por la Relatora Especial

[53] Comunicación (H) recibida por la Relatora Especial

[54] Comunicación (J) recibida por la Relatora Especial

[55] Comunicación (A) recibida por la Relatora Especial

[56] Comunicación (A) recibida por la Relatora Especial

[57] Comunicación (I) recibida por la Relatora Especial

[58] Comunicación (H) recibida por la Relatora Especial

[59] Comunicación (C) recibida por la Relatora Especial

a menudo se traduce al ámbito real [60]. Según la mencionada encuesta de Global Witness, el 84% de las personas defensoras del medio ambiente en África, América Latina y Asia que denunciaron ser blanco de ataques por su activismo creen que los daños que sufrieron en línea contribuyeron “directamente” o “en parte” a los daños sufridos fuera de línea [61].

4.2 FALTA DE MECANISMOS DE QUEJA EFICACES

En todas las comunicaciones recibidas, las organizaciones destacaron la falta de sistemas de queja ágiles, transparentes o impulsados por humanos cuando las personas defensoras de los derechos humanos tratan de denunciar abusos. La encuesta de Global Witness encontró que solo el 12% de las personas defensoras de los derechos humanos que denunciaron su abuso y acoso a las plataformas de redes sociales estuvieron satisfechas con la respuesta que recibieron de ellas [62]. Según una de las organizaciones que presentaron comunicaciones para este documento, las plataformas como TikTok y Telegram no responden de ninguna manera a peticiones de ayuda por parte de organizaciones especializadas en la región de **Centroamérica**. Es más, muchos procesos requieren fluidez en inglés y están gestionados por bots o respuestas automatizadas [63]. En **Bangladés**, una organización encontró que más del 90% de las denuncias de abuso que se hacían a Facebook en relación con acoso en línea de personas defensoras de derechos humanos o bien fueron ignoradas o bien resultaron en acciones de represalias contra la persona denunciante en lugar del infractor, incluyendo la suspensión de cuentas [64]. Una mujer DDH en **Liberia** destacó que, en algunos casos, la falta de respuesta de las plataformas de redes sociales a incidentes de ataques en línea contra mujeres DDH que promueven la participación política de las mujeres en el país había llevado a ataques físicos contra ellas [65]. Una organización que dedica tiempo y recursos a gestionar un servicio de soporte de seguridad digital para mujeres DDH informó que, eventualmente, la mayoría de las solicitudes que monitorea se resuelven. Sin embargo, el tiempo de respuesta de las plataformas de hasta 14 días, y la ocasional falta de respuesta o resolución negativa, contribuyen a poner en aún más peligro a DDH. Es más, es importante destacar que la mayoría de las personas DDH no tienen acceso a este tipo de apoyo y recursos sostenidos, lo que resulta en muchas menos resoluciones positivas, en comparación, en sus interacciones con las plataformas de redes sociales. También se expresaron preocupaciones sobre casos de cuentas falsas o acoso en chats, que son problemas comunes para las mujeres defensoras de derechos humanos en el **Oriente Medio y Norte de África** y pueden llevar a graves daños. Sin embargo, la organización informa que las

[60] Comunicaciones (B), (F), (H) y (J) recibidas por la Relatora Especial

[61] *Global Witness*, “Plataformas tóxicas, planeta destruido”, 16 de julio de 2025

[62] *Global Witness*, *Ibid*

[63] Comunicación (H) recibida por la Relatora Especial

[64] Comunicación (A) recibida por la Relatora Especial

[65] Comunicación (K) recibida por la Relatora Especial

plataformas fallan rutinariamente a la hora de resolver tales problemas, y se limitan a pedir a los usuarios que rellenen un formulario sin resultados tangibles [66].

5. TRATAMIENTO DESIGUAL DE USUARIOS Y CONTENIDOS

Varias organizaciones identificaron disparidades en cómo las plataformas de redes sociales tratan a las personas defensoras de derechos humanos y el contenido que publican, especialmente basado en su idioma y región [67].

5.1 DISPARIDADES DE IDIOMA

Un problema fundamental es la falta de una capacidad de idiomas adecuada en los equipos y las herramientas de moderación de contenidos. Esto se hace especialmente evidente en la moderación de cualquier idioma que no sea inglés, francés o español, en el que los contenidos dañinos o de odio a menudo pasan desapercibidos y desatendidos.

La investigación de una organización sobre el acoso en línea en **Bangladés** ha demostrado que las plataformas como Facebook no logran detectar y responder a contenidos abusivos en bengalí. A pesar de denuncias repetidas, las cuentas abusivas a menudo permanecen activas, y los términos de odio como “*shahbagi*” [68], usados de manera despectiva hacia personas defensoras de los derechos humanos y activistas, no se marcan ni eliminan [69]. Al mismo tiempo, la sobremoderación en otros idiomas crea una dinámica diferente pero igualmente dañina. Varias organizaciones señalaron que las publicaciones en idioma árabe, especialmente aquellas relativas a **Palestina, Siria** y temas más amplios de derechos regionales, están moderadas más agresivamente que las publicaciones en otros idiomas [70].

En general, los contenidos en idiomas diferentes al inglés, francés o español tienden a ser retirados o marcados para revisión de manera desproporcionada, ya que las plataformas

[66] Comunicación (C) recibida por la Relatora Especial

[67] Comunicaciones (A), (B), (C), (F) e (I) recibidas por la Relatora Especial

[68] Vocablo que originó en las protestas de 2013 de Shahbag, usado como término despectivo para referirse a individuos de izquierdas, liberales o laicos, a menudo para implicar islamofobia o apoyo a los juicios mediáticos y la justicia popular.

[69] Comunicación (A) recibida por la Relatora Especial

[70] Comunicaciones (A), (B) y (C) recibidas por la Relatora Especial

de redes sociales carecen de capacidad de moderación en idiomas locales [71]. Una organización de derechos humanos en **Camboya** expresó preocupaciones similares, señalando que las retiradas de contenido en idioma jemer no venían acompañadas de explicaciones, y que no existía ninguna infraestructura de apoyo para personas DDH que solo hablan jemer [72]. Otra organización informó que el urdu era un “punto ciego” para las plataformas en cuanto a discursos de odio y amenazas contra las personas defensoras de los derechos humanos en **Pakistán** [73]. En **Irán**, el proceso de Instagram de moderar contenidos en persa también ha expuesto deficiencias en la aplicación de las políticas de moderación de contenidos en idiomas no occidentales [74].

5.2 DISPARIDADES REGIONALES

Varias organizaciones también señalaron disparidades regionales importantes en cuanto a cómo las personas defensoras de los derechos humanos y sus contenidos son tratados en las plataformas de redes sociales [75]. Según investigaciones realizadas por una de las organizaciones que han presentado comunicaciones para este informe, a pesar de que el ritmo al que las personas DDH denuncian daños contra ellas está bastante igualado entre diferentes regiones globales, las personas defensoras **africanas** tenían una menor probabilidad de recibir una respuesta a su queja por parte de las plataformas que sus homólogos europeos (50% vs. 72% respectivamente) [76]. Otra organización identificó brechas similares en cuanto a cómo las plataformas abordan el acoso contra personas DDH en diferentes regiones [77].

6. DIFICULTADES AL INTERACTUAR CON LAS PLATAFORMAS DE REDES SOCIALES

6.1 RESULTADOS DECEPCIONANTES PARA LAS PERSONAS DEFENSORAS DE LOS DERECHOS HUMANOS

[71] Comunicaciones (A), (B), (C) y (E) recibidas por la Relatora Especial

[72] Comunicación (E) recibida por la Relatora Especial

[73] Comunicación (B) recibida por la Relatora Especial

[74] Article 19, “Iran: Meta must overhaul Persian-language content moderation on Instagram” (*Irán: Meta debe reformar la moderación de contenidos en lengua persa en Instagram*), junio de 2022 <https://www.article19.org/resources/iran-meta-persian-language-content-moderation-instagram/>

[75] Comunicaciones (A), (B), (F) e (I) recibidas por la Relatora Especial

[76] Comunicación (I) recibida por la Relatora Especial

[77] Comunicación (A) recibida por la Relatora Especial

A la luz de las mencionadas restricciones, violaciones de derechos y sesgos, las personas defensoras de los derechos humanos tienen un acceso muy limitado a recursos.

Las investigaciones realizadas por una organización muestran que, dada la ausencia de un apoyo dedicado por parte de la sociedad civil para acompañar a personas DDH en sus quejas a plataformas de redes sociales, la eficacia de sus respuestas continúa siendo extremadamente limitada. Las plataformas como Meta tienen normas comunitarias que a menudo no reconocen que muchos tipos de abusos en línea realmente infringen sus políticas, especialmente cuando el contenido dañino se enmarca como parte de un debate público o la libertad de expresión. Como resultado, muchas denuncias de acoso son ignoradas o rechazadas, dejando a los defensores de derechos humanos sin recurso significativo. Incluso cuando las plataformas sí actúan, la respuesta a menudo resulta lenta y no concuerda con la gravedad del abuso [78].

Otra organización identificó que los principales obstáculos para el acceso de las defensoras de los derechos humanos a recursos en caso de ataques digitales en **Mesoamérica** eran la falta de mecanismos adecuados, dificultades para acceder a ellas, especialmente para aquellas que no hablan inglés, los largos periodos de espera para obtener una respuesta, la falta de seguimiento posterior, y el hecho de que las respuestas a menudo estaban automatizadas y gestionadas por *bots*, lo cual hace que sea casi imposible obtener una respuesta humana [79]. Como resultado, algunas personas defensoras escogen no denunciar los abusos que sufren. El 65% de las mujeres defensoras de los derechos humanos que denunciaron ataques digitales a esta organización no los denunciaron a las plataformas en sí, principalmente debido a la ineficacia de sus mecanismos [80]. La investigación de otra organización acerca del acoso en línea en **Bangladés** mostró un patrón parecido; las personas defensoras de los derechos humanos que fueron entrevistadas explicaron que los mecanismos de denuncia de Facebook eran a menudo complicados, opacos y desalentadores, lo que llevaba a muchos a desilusionarse y dejar de perseguir quejas formales [81].

En algunos casos que incluyen amenazas directas o uso desautorizado de imágenes de las personas defensoras, las plataformas ocasionalmente han respondido eliminando las publicaciones o los perfiles infractores. Sin embargo, tales intervenciones son raras e inconsistentes. De igual manera, cuando se trata de apelaciones relativas a los bloqueos y suspensiones, es raro que haya una resolución positiva.

La opacidad de los procesos de quejas y apelaciones, la falta de transparencia, y la ausencia de interlocutores humanos contribuyen a que las personas defensoras de los

[78] Comunicación (A) recibida por la Relatora Especial

[79] Comunicación (H) recibida por la Relatora Especial

[80] Comunicación (H) recibida por la Relatora Especial

[81] Comunicación (A) recibida por la Relatora Especial

derechos humanos raramente reciban las respuestas y los recursos que buscan. Esto lleva a que muchas personas defensoras busquen apoyo de organizaciones de la sociedad civil externas para tratar de obtener resultados.

6.2 LA SOCIEDAD CIVIL COMO INTERMEDIARIA: UN MODELO DEFICIENTE E INSOSTENIBLE

Algunas organizaciones de derechos humanitarios bien establecidas mantienen canales directos de comunicación con plataformas de redes sociales a fin de marcar casos para revisar y extender el apoyo a las personas defensoras de los derechos humanos. Si bien estos canales han permitido unas resoluciones comparativamente más satisfactorias y unos procesos más rápidos, las organizaciones de la sociedad civil a menudo las encuentran poco fiables o consistentes [82]. Una organización asegura que la interacción con las plataformas de redes sociales en representación de personas defensoras es a menudo poco productiva, y no lleva a ningún cambio sistémico significativo [83]. De igual manera, otra dijo que los resultados nunca estaban garantizados: cuando trató de obtener la reactivación de la cuenta de una persona defensora de derechos humanos **kazaja** mediante el establecimiento de una línea de comunicación directa entre el defensor de los derechos humanos y Meta, no recibió ningún seguimiento, a pesar de intentos repetidos [84].

Incluso las iniciativas como el Programa Socio de Confianza (*Trusted Partner*) de Meta, creado para fomentar la colaboración entre la plataforma y las organizaciones de la sociedad civil, han dado resultados inconsistentes. Una organización señaló que, en las ocasiones en que el contenido de defensores de los derechos humanos estaba suprimido o marcado para su revisión, trataba de realizar seguimientos a través de estos canales de apoyo oficiales. Si bien las respuestas del canal de Socio de Confianza de Meta eran más frecuentes que las de X o Telegram, a menudo eran respuestas automatizadas que carecían de justificación sustancial. Aun cuando se proporcionaban explicaciones, raramente reconocían o comprendían el contexto más amplio de la situación, y el escalamiento no garantizaba ninguna resolución [85].

En general, las interacciones entre las organizaciones de la sociedad civil y las plataformas parecen estar marcadas por inconsistencia, falta de transparencia y un creciente desentendimiento por parte de las empresas. A pesar de esto, las ONG están haciendo lo posible por llenar los huecos que dejan las plataformas y cada vez más desempeñan roles intermediarios, documentando daños, interactuando con las plataformas, y abogando por

[82] Comunicaciones (A), (B), (C), (F) e (I) recibidas por la Relatora Especial

[83] Comunicación (I) recibida por la Relatora Especial

[84] Comunicación (B) recibida por la Relatora Especial

[85] Comunicación (A) recibida por la Relatora Especial

por los usuarios [86]. Sin embargo, esto no es un modelo sostenible, ya que asigna la carga de la protección y el apoyo a la sociedad civil, mientras persiste la falta de mecanismos accesibles y efectivos para las personas defensoras de los derechos humanos que están en situaciones de riesgo [87].

6-3 DETERIORO DE LOS COMPROMISOS DE LAS PLATAFORMAS CON LOS DERECHOS HUMANOS

En los últimos años, en una clara muestra de su compromiso deteriorado hacia los derechos humanos, las principales empresas de redes sociales han comenzado a reducir gradualmente su colaboración con las personas defensoras de los derechos humanos y las ONG.

A pesar de una larga historia de trabajar directamente con las plataformas de redes sociales, tanto informalmente como en calidad de miembro oficial de varios programas de socios de confianza, una organización declaró que la interacción era cada vez menos productiva, con unos tiempos de respuesta que aumentaron de manera importante en 2025. A veces se dejaban publicados contenidos dañinos hasta 10 a 15 días después de haber sido denunciados, incluso en casos de amenazas graves contra mujeres defensoras de derechos humanos en el **Oriente Medio y Norte de África**. También se informó que las plataformas de redes sociales eran mucho menos reactivas que antes ante la presión pública [88]. Otras organizaciones de la sociedad civil coincidieron en que la implicación de las plataformas en materia de derechos humanos ha caído de manera importante en los últimos tres años, mostrando cada vez menos interés en las aportaciones de la sociedad civil [89].

Estos cambios ocurrieron en paralelo a que las empresas de redes sociales recurrieran a despidos masivos de personal y equipos clave dedicados al fomento de la confianza, la seguridad de los usuarios, los derechos humanos y a contrarrestar la desinformación, y al desmantelamiento de instituciones y programas de asesoría centrados en estos temas.

[86] Comunicaciones (A), (B) y (C) recibidas por la Relatora Especial

[87] Comunicación (A) recibida por la Relatora Especial

[88] Comunicación (C) recibida por la Relatora Especial

[89] Comunicaciones (I) y (J) recibidas por la Relatora Especial

Las plataformas afectadas incluyen X [90], Meta [91], YouTube [92] y TikTok [93]. Tales cambios estructurales han debilitado de manera sustancial cualquier vía de diálogo que previamente tenían las personas defensoras de los derechos humanos y las organizaciones de la sociedad civil, y han impactado la capacidad de las plataformas para abordar los casos individuales y los problemas estructurales [94].

Simultáneamente, muchas plataformas de redes sociales dieron marcha atrás a políticas fundamentales que contribuían a la protección de los derechos de usuarios. Meta, X y YouTube han debilitado todas las protecciones de privacidad de los usuarios para permitir el entrenamiento de sus herramientas IA. Meta ha reducido la comprobación-de datos de terceros a favor de una moderación “comunitaria” o automatizada, lo cual genera preocupaciones sobre la divulgación de desinformación. X y YouTube también han dado marcha atrás a sus políticas de desinformación sobre elecciones [95].

Si bien Meta ha publicado informes anuales sobre derechos humanos desde 2022, detallando cómo la empresa identifica y mitiga los riesgos de derechos humanos en sus plataformas, su última edición [96] no incluye referencias a los mencionados retos de manera sustancial. Otras plataformas, como X, TikTok o YouTube, no publican informes específicos e independientes sobre los derechos humanos.

[90] Amnistía Internacional, “Global: La decisión de Twitter de disolver el consejo de seguridad hace peligrar el bienestar de sus usuarios y usuarias”, diciembre de 2022 <https://www.amnesty.org/es/latest/news/2022/12/global-twitthers-decision-to-disband-safety-council-threatens-wellbeing-of-users/>; Electronic Frontier Foundation, “Monetization, Not Human Rights or Vulnerable Communities, Matter Most at Twitter Under Musk” (*En el Twitter de Musk, es la monetización lo que importa, no los derechos humanos ni las comunidades vulnerables*), noviembre de 2022 <https://www.eff.org/deeplinks/2022/11/twitthers-monetizable-users-not-human-rights-matter-most-under-musks-rein-leaving>; Comunicación (F) recibida por la Relatora Especial

[91] Free Press, “Big Tech Backslide: How Social-Media Rollbacks Endanger Democracy Ahead of the 2024 Elections” (*Reculada Big Tech: Cómo los retrocesos de las redes sociales ponen en peligro a la democracia antes de las elecciones de 2024*), diciembre de 2023 <https://www.freepress.net/big-tech-backslide-how-social-media-rollbacks-endanger-democracy-ahead-2024-elections>; Comunicación (F) recibida por la Relatora Especial

[92] Free Press, *ibid*

[93] Business and Human Rights Centre, “Germany: TikTok content moderators strike over mass layoffs & plans to replace them with AI” (*Alemania: huelga de los moderadores de contenidos de TikTok en protesta contra los despidos masivos y los planes para reemplazarlos con IA*), agosto de 2025 <https://www.business-humanrights.org/en/latest-news/deutschland-content-moderatorinnen-protestieren-gegen-stellenabbau-geplante-ersetzung-durch-ki-algorithmus/>

[94] Comunicación (C) recibida por la Relatora Especial

[95] Free Press, *ibid*

[96] Meta, “Informe sobre derechos humanos 2024”, diciembre de 2025 https://humanrights.fb.com/wp-content/uploads/2026/01/2024-Meta-Human-Rights-Report_es_LA.pdf

7. CONCLUSIÓN: IMPACTO SOBRE LA PRESENCIA EN LÍNEA Y LA ESTRATEGIA ORGANIZACIONAL DE LAS PERSONAS DEFENSORAS

El impacto acumulativo de las formas de represión digital mencionadas se traduce en un espacio cívico digital menguante, en el que las personas defensoras de los derechos humanos están siendo gradualmente silenciados en línea.

Al enfrentarse a restricciones de visibilidad, acoso o vigilancia a través de redes sociales, y al no tener un apoyo eficaz o adecuado por parte de las plataformas para remediar tales violaciones, un gran número de personas defensoras de los derechos humanos dicen tener miedo y ansiedad por ellas mismas y sus comunidades, como se aprecia claramente en las comunicaciones recibidas para este informe [97] y en las interacciones de la Relatora Especial con las personas defensoras de derechos humanos. Como resultado, pueden escoger reducir su visibilidad, evitar temas sensibles o abandonar las plataformas por completo.

Las mujeres defensoras de los derechos humanos en **El Salvador** se han retirado del discurso público debido al acoso digital sostenido que sufren [98]. En **Pakistán**, un estudio de 2023 realizado por Media Matters for Democracy encontró que 8 de cada 10 mujeres periodistas se estaban autocensurando en línea por miedo a la violencia digital y física [99]. En **Camboya**, una organización constató que un tercio de periodistas ahora evita publicar contenidos que sean críticos de las autoridades en las redes sociales [100]. En otros lugares de **Asia**, la autocensura sobre temas políticamente sensibles sigue siendo prevalente [101]. Casi la mitad de las personas defensoras de la tierra y el medio ambiente que fueron encuestadas por Global Witness afirmaron que el abuso que habían sufrido en línea había llevado a una pérdida de productividad y una capacidad reducida para realizar campañas sobre estos temas [102].

[97] Comunicaciones (A), (H), (I) y (J) recibidas por la Relatora Especial

[98] Comunicación (H) recibida por la Relatora Especial

[99] Comunicación (J) recibida por la Relatora Especial

[100] Comunicación (E) recibida por la Relatora Especial

[101] Comunicación (F) recibida por la Relatora Especial

[102] Global Witness, *ibid.*

El impacto general es una menor presencia digital y una participación pública reducida por parte de las personas defensoras de los derechos humanos, en especial aquellas que pertenecen a comunidades marginadas y aquellas que intervienen en asuntos delicados. Esto sienta las bases para un ecosistema en línea que refuerza las dinámicas de poder existentes, dejando a las comunidades que ya están sobreexpuestas a daños en riesgo de ser completamente borradas de los espacios digitales.

Algunas personas y grupos defensores de los derechos humanos están explorando plataformas alternativas, más autónomas, que priorizan la seguridad y la privacidad sobre el alcance. Otras han ajustado su estrategia en cuanto a presencia de redes sociales, decidiendo limitar sus huellas digitales, usar pseudónimos o lenguaje codificado y evitar publicar en tiempo real, entre otros [103]. Sin embargo, para la mayoría de las personas defensoras de los derechos humanos, las principales plataformas de redes sociales siguen siendo esenciales. Las plataformas más nuevas aún tienen un alcance muy limitado y un número mucho menor de usuarios activos, y el ecosistema digital actual sigue estando dominado por un puñado de grandes plataformas de redes sociales con fines de lucro. Esto lleva a que las personas defensoras de derechos humanos tengan que encontrar un equilibrio entre la visibilidad y la seguridad. Varias organizaciones nos comentaron la dificultad que supone para las personas defensoras salirse enteramente de estas plataformas sin arriesgarse a sacrificar el impacto de su labor, y la falta de alternativas seguras que lleva a la exposición continuada a riesgos [104].

[103] Comunicaciones (A), (C), (F) e (I) recibidas por la Relatora Especial

[104] Comunicaciones (A), (B) y (C) recibidas por la Relatora Especial

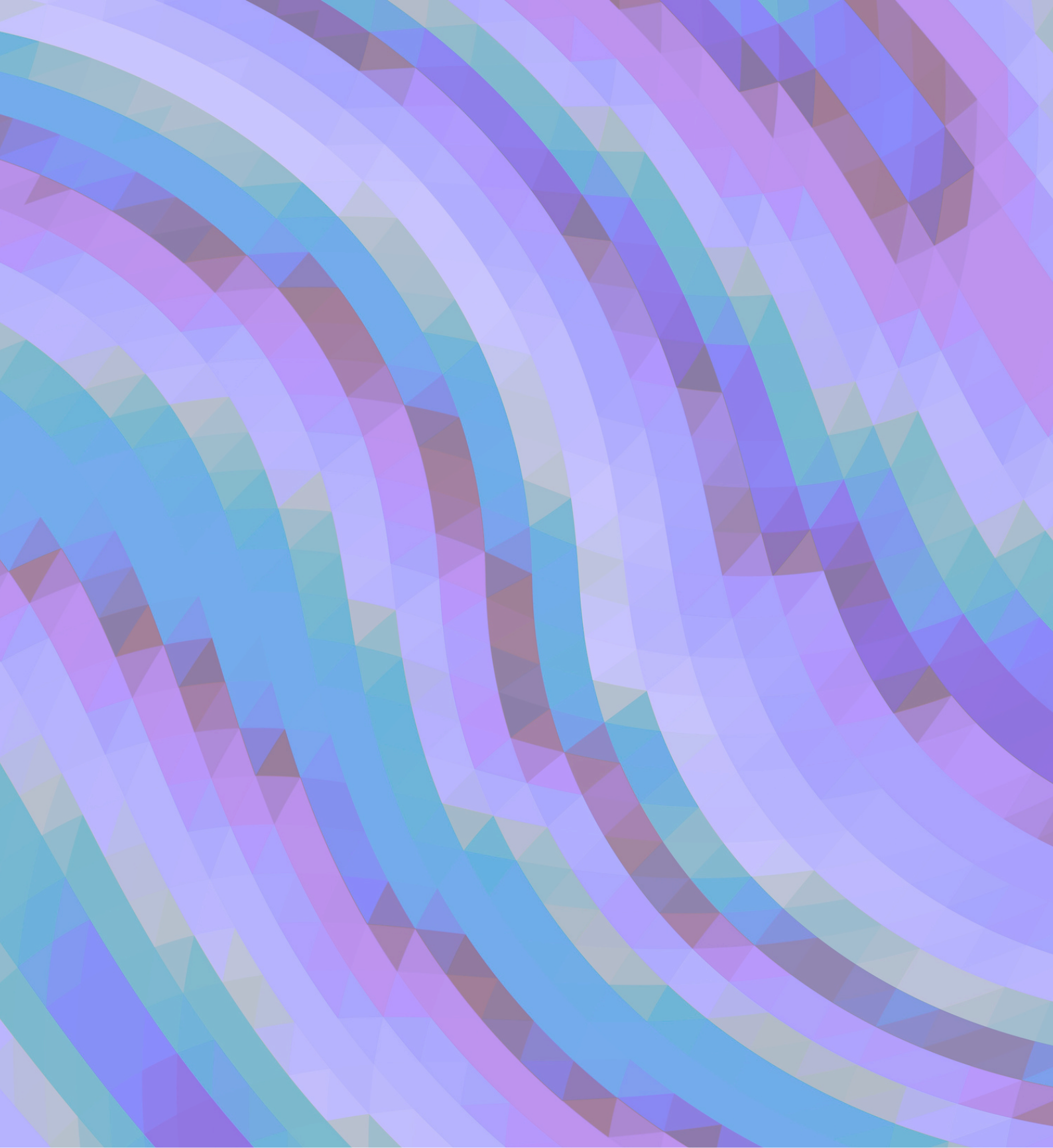
RECOMENDACIONES

LAS EMPRESAS QUE OPERAN LAS PLATAFORMAS DE REDES SOCIALES DEBERÍAN:

- Asegurar que **las políticas de moderación y gestión de contenidos** estén estrictamente alineadas con el derecho internacional de los derechos humanos y no menoscaben el derecho a defender los derechos humanos;
- Abstenerse de desarrollar y usar **IA** en las prácticas de moderación y eliminación de contenidos sin procesos de comprobación y verificación de datos humanos y externos;
- Mejorar la **transparencia** sobre las decisiones y recursos en la moderación de contenidos;
- Mejorar las **herramientas de privacidad y seguridad**, incluyendo un mayor nivel de cifrado, denuncias anónimas y salvaguardias contra el doxing y la vigilancia;
- Revisar la **eficacia de los mecanismos de denuncia** para los abusos en línea, e invertir recursos en **mecanismos de apoyo humanos** fiables y accesibles para proporcionar **asistencia** a las personas defensoras de los derechos humanos que lo necesiten, con equipos formados en los idiomas, las culturas y los contextos políticos pertinentes;
- Asegurar tiempos de respuesta más rápidos, especialmente con **socios de confianza y en casos urgentes**;
- Invertir en herramientas y protocolos de moderación **sensibles al género**, e incorporar en las directrices de la comunidad secciones específicas sobre acoso y abuso basados en género;
- Implementar **auditorías de algoritmos** externas para limitar la amplificación de la desinformación y contenidos odiosos o abusivos, y comprometerse a un diseño inclusivo que no margine los contenidos sobre derechos humanos;
- Oponerse a cumplir con **peticiones estatales de censura** que estén en conflicto con las normas internacionales de derechos humanos;
- Registrar públicamente y **publicar peticiones estatales de retirada** y las respuestas de las plataformas;
- Revertir los **retrocesos recientes en las políticas de discursos de odio, desinformación y privacidad** y en la contratación de personal de Confianza y Seguridad y Derechos Humanos;
- Establecer y expandir **canales de comunicación con personas defensoras de los derechos humanos y la sociedad civil**, especialmente en regiones en las que no existen mecanismos estructurados, y asegurar que tengan voz en el diseño de políticas y sistemas de apoyo que funcionen para ellos.

LOS ESTADOS DEBERÍAN:

- Asegurar que las empresas de redes sociales respeten la libertad de expresión y **respeten las normas de derechos humanos**;
- Contemplar **soluciones reglamentarias** para abordar el poder de las plataformas dominantes;
- Reconocer la violencia digital como parte de un continuo más amplio de ataques contra las personas DDH;
- Investigar y enjuiciar la violencia en línea contra las personas DDH;
- Adoptar **marcos legales y de políticas** que responsabilicen a las plataformas de redes sociales por su papel en la creación de un ambiente propicio para el acoso en línea, especialmente cuando tales acciones amenazan la seguridad de las personas DDH;
- Abstenerse de coaccionar a las plataformas de redes sociales para implementar censura o vigilancia.



HRC-SR-DEFENDERS@UN.ORG
WWW.SRDEFENDERS.ORG



RELATORA ESPECIAL DE
LA ONU SOBRE LOS
DEFENSORES DE
LOS DERECHOS
HUMANOS

SOLIDARIDAD Y ESPERANZA