

Informe de políticas de
Nuestra Agenda Común 8

**Integridad
de la información
en las plataformas
digitales**

JUNIO 2023



**Naciones
Unidas**



Prólogo

INTRODUCCIÓN

Los retos a los que nos enfrentamos solo pueden afrontarse mediante una mayor cooperación internacional. La Cumbre del Futuro de 2024 brinda la oportunidad de acordar soluciones multilaterales para un mañana mejor, fortaleciendo la gobernanza global por el bien de las generaciones presentes y venideras (resolución [76/307](#) de la Asamblea General). En mi calidad de Secretario General, he sido invitado a realizar aportaciones a los preparativos de la Cumbre en forma de recomendaciones orientadas a la acción, a partir de las propuestas presentadas en mi informe titulado “Nuestra Agenda Común” ([A/75/982](#)), que era a su vez una respuesta a la declaración sobre la conmemoración del 75º aniversario de las Naciones Unidas (resolución [75/1](#) de la Asamblea General). El presente informe es una de esas aportaciones. En él se desarrollan las ideas propuestas por primera vez en “Nuestra Agenda Común”, teniendo en cuenta las orientaciones posteriores proporcionadas por los Estados Miembros y los resultados de más de un año de consultas intergubernamentales y con múltiples partes interesadas, y se basa en los propósitos y principios de la Carta de las Naciones Unidas, la Declaración Universal de Derechos Humanos y otros instrumentos internacionales.

OBJETIVO DEL INFORME DE POLÍTICAS

El presente informe de políticas se centra en cómo las amenazas a la integridad de la información repercuten en el progreso de las cuestiones globales, nacionales y locales. En Nuestra Agenda Común, defendí un consenso común y empírico en torno a los hechos, la ciencia y el conocimiento. A tal fin, aquí se describen someramente los posibles principios de un código de conducta que ayude a orientar a los Estados Miembros, las plataformas digitales y otras partes interesadas en sus esfuerzos por hacer que el espacio digital sea más inclusivo y seguro para todos, defendiendo al mismo tiempo a capa y espada el derecho a la libertad de opinión y de expresión y el derecho de acceso a la información. El código de conducta para la integridad de la información en las plataformas digitales se está elaborando en el contexto de los preparativos de la Cumbre del Futuro. Mi esperanza es que sea una norma de referencia que oriente las medidas destinadas a reforzar la integridad de la información.

Las plataformas digitales son herramientas cruciales que han transformado las interacciones sociales, culturales y políticas en todo el mundo. En todo el mundo, ponen en contacto a ciudadanos preocupados por los asuntos importantes. Las plataformas ayudan a las Naciones Unidas a informar e implicar directamente a las personas en nuestra lucha por la paz, la dignidad y la igualdad en un planeta sano. Han dado esperanza a la gente en tiempos de crisis y lucha, han amplificado voces que antes no se escuchaban y han insuflado vida a los movimientos mundiales.

Sin embargo, estas mismas plataformas también han revelado un lado más oscuro del ecosistema digital. Han permitido la rápida propagación de la mentira y el odio, causando un daño real a escala mundial. El optimismo sobre el potencial de las redes sociales para conectar e implicar a la gente se ha visto frenado por la información errónea y la desinformación y el discurso de odio, que han saltado de los márgenes del espacio digital a la palestra. No se debe subestimar el peligro. El discurso de odio y la desinformación a través de las redes sociales pueden provocar violencia y muertes.¹ La capacidad de generar desinformación a gran escala y atentar contra los hechos

establecidos científicamente es un riesgo existencial para la humanidad ([A/75/982](#), párr. 26) y pone en peligro las instituciones democráticas y los derechos humanos fundamentales. Estos riesgos se han intensificado aún más debido a los rápidos avances de la tecnología, como la inteligencia artificial generativa. En todo el mundo, las Naciones Unidas están monitoreando la amenaza que pueden suponer la información errónea y la desinformación y el discurso de odio para el progreso hacia los Objetivos de Desarrollo Sostenible. Ha quedado claro que no podemos seguir como hasta ahora.

EN LA PRÁCTICA,

NO ES FÁCIL

DISTINGUIR ENTRE

**LA INFORMACIÓN ERRÓNEA Y
LA DESINFORMACIÓN**

¿Qué es la integridad de la información?

La integridad de la información se refiere a la exactitud, la coherencia y la fiabilidad de la información. Está amenazada por la desinformación, la información errónea y el discurso de odio. Aunque no existen definiciones aceptadas universalmente de estos términos, las entidades de las Naciones Unidas han elaborado definiciones de trabajo.

La Relatora Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión define la desinformación como “información falsa que se difunde intencionadamente para causar un grave perjuicio social”.² La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) define la desinformación como contenidos falsos o engañosos que pueden provocar un daño específico, independientemente de las motivaciones, la conciencia o los comportamientos.³

A efectos del presente informe de políticas, la diferencia entre la información errónea y la desinformación radica en la intención. La desinformación es información que no solo es inexacta, sino que además pretende engañar y se difunde para causar daño. Puede ser difundida por agentes estatales o no estatales en múltiples contextos, incluso durante los conflictos armados, y puede afectar a todos los ámbitos del desarrollo, desde la paz y la seguridad hasta los derechos humanos, la salud pública, la ayuda humanitaria y la acción climática.

La información errónea se refiere a la difusión involuntaria de información inexacta compartida de buena fe por quienes no son conscientes de que están transmitiendo falsedades. Puede tener sus raíces en la desinformación, ya que las mentiras deliberadas y los mensajes engañosos se convierten en armas a lo largo del tiempo, se introducen en el discurso público y se transmiten involuntariamente.⁴ En la práctica, no es fácil distinguir entre la información errónea y la desinformación.

Según la definición de trabajo recogida en la Estrategia y Plan de Acción de las Naciones Unidas para la Lucha contra el Discurso de Odio, el discurso de odio es “cualquier forma de comunicación de palabra, por escrito o a través del comportamiento, que sea un ataque o utilice lenguaje peyorativo o



discriminatorio en relación con una persona o un grupo sobre la base de quiénes son o, en otras palabras, en razón de su religión, origen étnico, nacionalidad, raza, color, ascendencia, género u otro factor de identidad”.⁵

La información errónea y la desinformación y el discurso de odio son fenómenos relacionados pero distintos, con algún solapamiento y con diferencias en la forma de detectarlos, mitigarlos y abordarlos. Los tres contaminan el ecosistema de la información y son una amenaza para el progreso de la humanidad.⁶

Las amenazas a la integridad de la información no son nuevas. Las falsedades y el odio se han difundido durante mucho tiempo para obtener beneficios políticos o económicos. Sin embargo, en la era digital estas operaciones pueden llegar a una escala antes impensable. Las potentes herramientas de comunicación pueden difundir contenido instantáneamente por todo el planeta, creando un problema tan extendido que las propias plataformas en línea son a veces incapaces de captar toda su magnitud. La falta de definiciones de estos términos acordadas a nivel gubernamental no debe traducirse en inercia. Debemos hacer todo lo posible para frenar los daños que causan.

Integridad de la información y plataformas digitales

Las plataformas digitales deben ser parte integrante de la defensa de la integridad de la información. Aunque algunos medios de comunicación tradicionales también pueden ser fuentes de información errónea y desinformación, la velocidad, el volumen y la viralidad de su difusión a través de los canales digitales justifican la necesidad de dar una respuesta urgente y adaptada. A efectos del presente informe, el término "plataforma digital" se refiere a un servicio digital que facilita las interacciones entre dos o más usuarios y abarca una amplia gama de actividades, desde las redes sociales y los motores de búsqueda hasta las aplicaciones de mensajería. Normalmente, recopilan datos sobre los usuarios y sus interacciones.⁷

La información errónea y la desinformación proceden de muchas fuentes, con diversas motivaciones, que en general logran permanecer en el anonimato. Ha habido campañas coordinadas de desinformación protagonizadas por agentes estatales y no estatales que han explotado sistemas digitales defectuosos para promover historias perjudiciales, con graves repercusiones.

Muchos Estados han puesto en marcha iniciativas para regular las plataformas digitales, en los últimos cuatro años se han aprobado o estudiado al menos 70 leyes de ese tipo.⁸ En esencia, los enfoques legislativos suelen implicar un ámbito limitado de reparaciones para definir y eliminar el contenido nocivo. Al centrarse en esto último, algunos Estados han introducido una legislación defectuosa y excesivamente amplia que ha acabado silenciando la "protección de la libertad de expresión", permitida por el derecho internacional. Otras respuestas, como los apagones generales de Internet y las prohibiciones dictadas

contra algunas plataformas, pueden carecer de base jurídica e infringir los derechos humanos.

Muchos Estados y personalidades políticas han utilizado la supuesta preocupación por la integridad de la información como pretexto para restringir el acceso a la información, desacreditar y restringir la labor informativa, y atacar a periodistas y opositores.⁹ Los agentes estatales también han presionado a las plataformas para que cumplan sus órdenes con el pretexto de combatir la información errónea y la desinformación.¹⁰ Los expertos en libertad de expresión han subrayado que los agentes estatales tienen un deber especial en este contexto y "no deben fabricar, patrocinar, fomentar o promover información falsa" (A/77/287, párr. 45).

Incluso si existe un propósito legítimo de interés público, los riesgos inherentes a la regulación de la expresión requieren un enfoque cuidadosamente adaptado que cumpla con los requisitos de legalidad, necesidad y proporcionalidad en virtud de la legislación de derechos humanos (ibid., párr. 42).

La desinformación también es un gran negocio. Tanto las empresas de relaciones públicas "oscuras" como las convencionales, contratadas por Estados, personalidades políticas y el sector privado, son fuentes clave de contenidos falsos y engañosos.¹¹ Una táctica, entre otras, ha consistido en publicar contenido en copias falsas de sitios de noticias para que parezca que los artículos proceden de fuentes legítimas.¹² Este oscuro negocio es extremadamente difícil de rastrear e investigar, por lo que se desconoce la verdadera magnitud del problema. También las personas difunden afirmaciones falsas para vender productos o servicios con ánimo de lucro, a menudo

dirigidos a grupos vulnerables en tiempos de crisis o inseguridad.

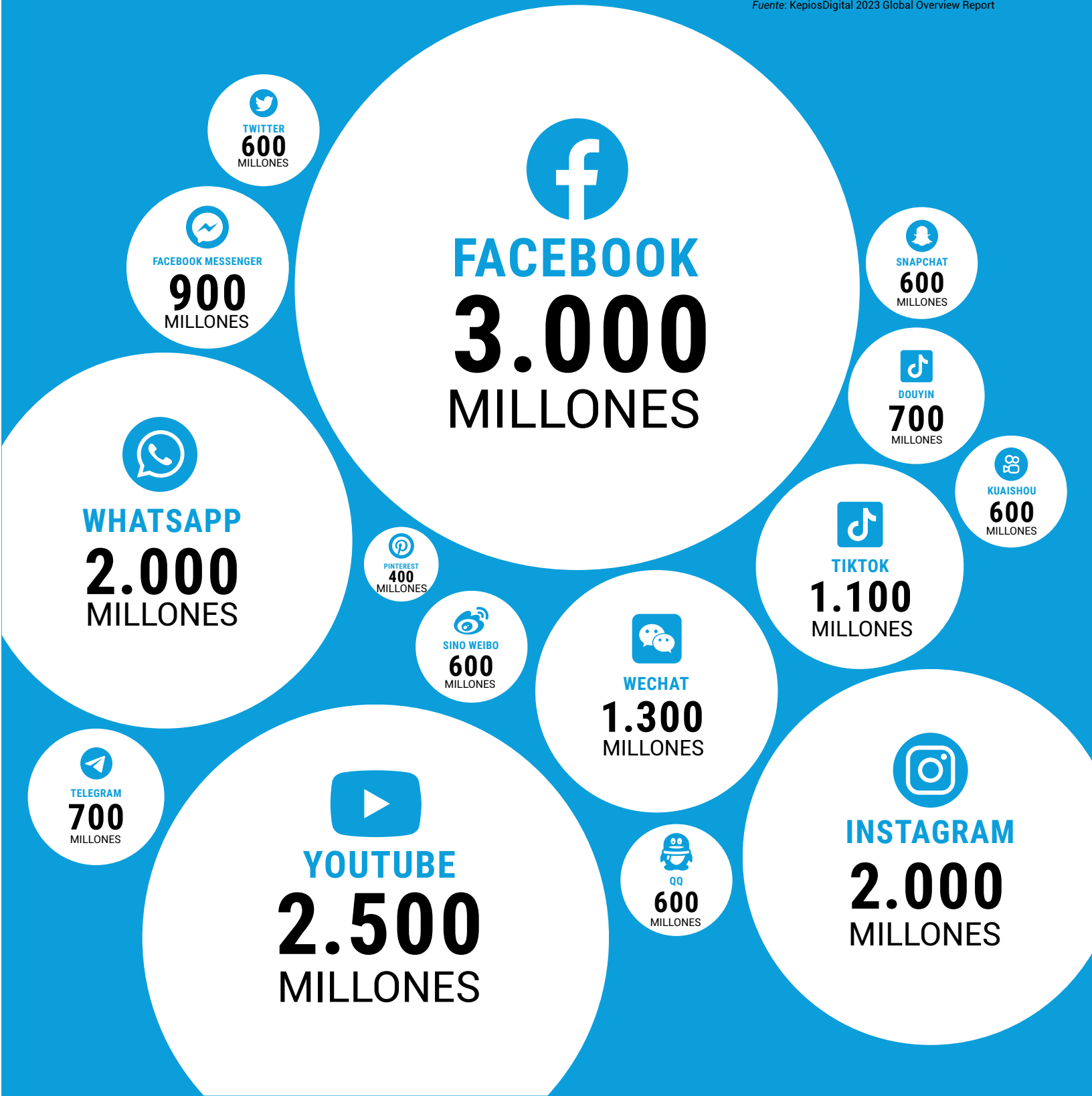
Un enfoque dominante en los modelos de negocio actuales de la mayoría de las plataformas digitales gira en torno a la "economía de la atención". Los algoritmos están diseñados para dar prioridad a los contenidos que mantienen la atención de los usuarios, maximizando así la interacción y los ingresos por publicidad. El contenido inexacto y odioso diseñado para polarizar a los usuarios y provocar emociones fuertes es a menudo el que genera más participación, con el resultado de que se sabe que los algoritmos recompensan y amplifican la información errónea y la desinformación y el discurso de odio.¹³

Ante el descenso de los ingresos publicitarios, las plataformas digitales buscan vías alternativas para obtener beneficios más allá de la economía de la atención. Por ejemplo, los sistemas de verificación de pago, por los que los usuarios pueden comprar un sello de aprobación que antes se utilizaba para denotar autenticidad, han suscitado serias preocupaciones por la integridad de la información, dado el potencial de abuso por parte de los agentes de la desinformación.¹⁴

MILES DE MILLONES

DE USUARIOS EN LAS REDES SOCIALES

Fuente: KepiosDigital 2023 Global Overview Report



¿Cuál es el marco jurídico internacional pertinente?

La promoción de la integridad de la información debe basarse plenamente en las normas y estándares internacionales pertinentes, incluidas las normas de derechos humanos y los principios de soberanía y no intervención en los asuntos internos de otros Estados. El derecho humano a la libertad de expresión no se limita a la información recibida favorablemente (A/77/287, párr. 13). Vinculada a la libertad de expresión, la libertad de información es en sí misma un derecho. La Asamblea General ha manifestado: "La libertad de información es un derecho humano fundamental y piedra de toque de todas las libertades a las cuales están consagradas las Naciones Unidas" (ibid., párr. 14). En agosto de 2022, transmití a la Asamblea General un informe titulado "Contrarrestar la desinformación para promover y proteger los derechos humanos y las libertades fundamentales".¹⁵ En él, expuse la legislación internacional sobre derechos humanos que se aplica a la desinformación, incluida la Declaración Universal de Derechos Humanos y el Pacto Internacional de Derechos Civiles y Políticos. En virtud de estos instrumentos jurídicos internacionales, toda persona tiene derecho a la libertad de expresión.¹⁶

El artículo 19 de la Declaración Universal de Derechos Humanos y el artículo 19 2) del Pacto protegen el derecho a la libertad de expresión, que incluye la libertad de buscar, recibir y difundir informaciones e ideas de todo tipo, sin limitación de fronteras, y por cualquier medio de expresión. La libertad de expresión y el acceso a la información pueden estar sujetos a ciertas restricciones que cumplan criterios específicos descritos en el artículo 19 3) del Pacto.¹⁷ Los Estados no pueden añadir motivos adicionales ni restringir la expresión más allá de lo que permite el derecho internacional.

El Plan de Acción de Rabat sobre la prohibición de la apología del odio nacional, racial o religioso que constituye incitación a la discriminación, la hostilidad o la violencia, aprobado en 2012, proporciona orientación jurídica y política práctica a los Estados sobre la mejor manera de aplicar el artículo 20 2) del Pacto y el artículo 4 de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, que prohíben ciertas formas de discurso de odio. El Plan de Acción de Rabat ya ha sido utilizado por los Estados Miembros en diferentes contextos.¹⁸

El discurso de odio ha sido precursor de crímenes atroces, incluido el genocidio. La Convención para la Prevención y la Sanción del Delito de Genocidio de 1948 prohíbe la "incitación directa y pública a cometer genocidio".

En su resolución [76/227](#), adoptada en 2021, la Asamblea General puso de relieve que todos los tipos de desinformación pueden afectar negativamente al disfrute de los derechos humanos y las libertades fundamentales, así como a la consecución de los Objetivos de Desarrollo Sostenible. Asimismo, en su resolución [49/21](#), adoptada en 2022, el Consejo de Derechos Humanos afirmó que la desinformación puede afectar negativamente al disfrute y la realización de todos los derechos humanos.

EL DISCURSO DE ODIO

HA SIDO PRECURSOR

DE CRÍMENES ATROCES,

INCLUIDO

EL GENOCIDIO



El 75 %

**DEL PERSONAL DE MANTENIMIENTO
DE LA PAZ DE LA ONU AFIRMÓ QUE
LA INFORMACIÓN ERRÓNEA Y
LA DESINFORMACIÓN**

AFECTABAN A

SU

SEGURIDAD

ENCUESTA INTERNA – ONU, 2022

¿Qué daños causan la desinformación y la incitación al odio en Internet?

La desinformación y la incitación al odio en Internet preocupan seriamente a la opinión pública mundial. Un estudio realizado a partir de datos de encuestados de 142 países reveló que al 58,5 % de los usuarios habituales de Internet y de las redes sociales de todo el mundo les preocupa encontrarse con información errónea en línea, y que los jóvenes y las personas con bajos ingresos¹⁹ se sienten mucho más vulnerables. Los jóvenes de hoy son nativos digitales con más probabilidades de estar conectados en línea que el resto de la población, lo que los convierte en la generación más conectada²⁰ digitalmente de la historia. En todo el mundo, cada medio segundo un niño se conecta por primera vez a Internet, lo que le expone a la incitación al odio y a otros daños que, en algunos casos, afectan a su salud mental.²¹

Las repercusiones de la información errónea y la desinformación y el discurso de odio en Internet se dejan sentir en todo el mundo, incluso en los ámbitos de la salud, la acción climática, la democracia y las elecciones, la igualdad de género, la seguridad y la respuesta humanitaria. Según una encuesta realizada en 2021, el 75 % de las oficinas del Programa de las Naciones Unidas para el Desarrollo en los países señalaron que la contaminación informativa era un problema importante.

UN ESTUDIO REALIZADO A PARTIR DE DATOS DE ENCUESTADOS

DE 142 PAÍSES REVELÓ QUE

AL 58,5 %

DE LOS USUARIOS HABITUALES DE INTERNET Y DE LAS REDES SOCIALES DE TODO EL MUNDO LES PREOCUPA ENCONTRARSE CON

INFORMACIÓN ERRÓNEA EN LÍNEA

Tiene graves consecuencias para la confianza, la seguridad, la democracia y el desarrollo sostenible, según se desprende de un reciente estudio encargado por la UNESCO de más de 800 documentos académicos, de la sociedad civil, periódicos y empresariales.²²

La información errónea y la desinformación pueden ser peligrosas y hasta mortales, especialmente en tiempos de crisis, emergencia o conflicto. Durante la pandemia de enfermedad por coronavirus (COVID-19), empezó a circular por Internet una avalancha de información errónea y desinformación sobre el virus, las medidas de salud pública y las vacunas.²³ Ciertos actores explotaron la confusión para sus propios objetivos, con campañas antivacunas que conducían a los usuarios a sitios que vendían curas o medidas preventivas falsas.²⁴ Muchas víctimas de COVID-19 se negaron a vacunarse o a tomar precauciones sanitarias básicas tras verse expuestas a información errónea y desinformación en Internet.²⁵

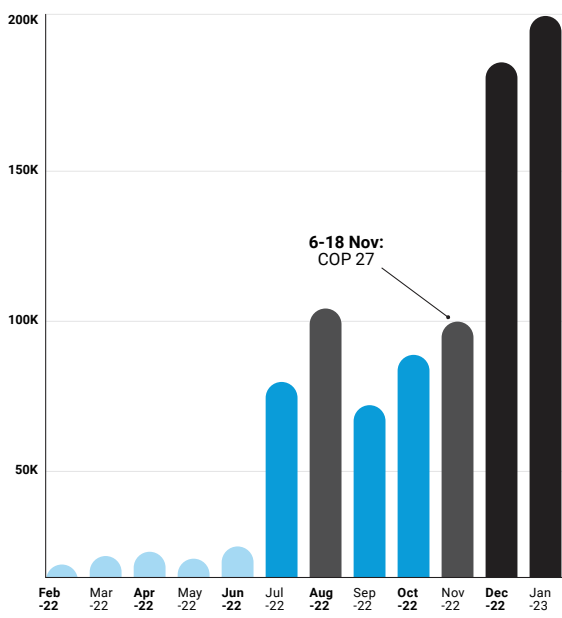
La desinformación también puede ser letal en contextos sociales y políticos ya de por sí volátiles. En un informe de 2022, la Relatora Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión examinó el impacto de la información convertida en arma a la hora de sembrar la confusión, alimentar el odio, incitar a la violencia y prolongar los conflictos.²⁶ En otro informe publicado en 2022, se constató que la desinformación puede “incluir el fanatismo y el discurso de odio contra las minorías, las mujeres y los llamados “otros”, con lo que supone de amenaza no solo para las personas directamente afectadas, sino también para la inclusión y la cohesión social. Puede amplificar las tensiones y divisiones en tiempos de emergencia, crisis, momentos políticos clave o conflictos armados”.²⁷

Algunos de los peores impactos de los daños en línea se producen en contextos desatendidos por las plataformas digitales, incluso allí donde éstas gozan de altos índices de penetración. Los países en conflicto, o en situaciones volátiles por otros motivos, que suelen ser mercados menos lucrativos, no han recibido recursos suficientes para moderar contenidos o ayudar a los usuarios. Aunque los medios de comunicación tradicionales siguen siendo la fuente más común de noticias para la mayoría de los habitantes de las zonas de conflicto, el odio difundido en las plataformas digitales también ha desencadenado y avivado la violencia.²⁸ Algunas plataformas digitales se han enfrentado a críticas por su papel en determinados conflictos, como la guerra en Ucrania.²⁹

Del mismo modo, la información errónea y la desinformación sobre la emergencia climática está retrasando la adopción de medidas urgentes para garantizar un futuro habitable al planeta. La información errónea y la desinformación en cuanto al clima pueden entenderse como contenido falso o engañoso que socava la base científica de la existencia del cambio climático antropógeno,

sus causas y sus consecuencias. Hay campañas coordinadas que tratan de negar o minimizar el consenso científico del Grupo Intergubernamental de Expertos sobre el Cambio Climático o desviar la atención de él, y desbaratar las acciones urgentes necesarias para cumplir los objetivos del Acuerdo de París de 2015. Una pequeña pero ruidosa minoría de negacionistas climáticos³⁰ sigue rechazando la posición de consenso y goza de una presencia destacada en algunas plataformas digitales. Por ejemplo, en 2022, varias organizaciones de la sociedad civil hicieron simulaciones aleatorias que revelaron que el algoritmo de Facebook recomendaba contenido negacionista en detrimento de la climatología.³¹ En Twitter, la etiqueta #climatescam pasó de menos de 2.700 apariciones al mes en la primera mitad de 2022 a 80.000 en julio y 199.000 en enero de 2023. También aparecía en la plataforma entre los primeros resultados al buscar la palabra “clima”.³² En febrero de 2022, el Grupo Intergubernamental de Expertos sobre el Cambio Climático denunció por primera vez la desinformación climática, afirmando que un debilitamiento deliberado de la ciencia estaba contribuyendo a la aparición de percepciones erróneas del consenso científico, incertidumbre, desprecio del riesgo y la urgencia, y disensión”.³³

FIGURA I
MENCIONES MENSUALES DE #CLIMATESCAM EN TWITTER



Fuente: Departamento de Comunicación Global, con datos de Talkwalker

Algunas empresas de combustibles fósiles suelen desplegar una estrategia de “ecoimpostura”, engañando al público para que crea que una empresa o entidad hace más por proteger el medio ambiente y menos por dañarlo de lo que en realidad hace. Las empresas no actúan solas. Detrás de los esfuerzos por confundir al público y desviar la atención de la responsabilidad de la industria de los combustibles fósiles hay proveedores de servicios de publicidad y relaciones públicas, empresas de tecnología publicitaria, medios de comunicación y plataformas digitales.³⁴ Las empresas de publicidad y relaciones públicas que crean contenido de ecoimpostura y los terceros que los distribuyen ganan miles de millones con estos esfuerzos por proteger a la industria de los combustibles fósiles del escrutinio y la rendición de cuentas. Las empresas de relaciones públicas han realizado cientos de campañas para empresas de carbón, petróleo y gas.³⁵

EL 70% DEL PERSONAL DE MANTENIMIENTO DE LA PAZ DE LA ONU AFIRMÓ QUE LA INFORMACIÓN ERRÓNEA Y LA DESINFORMACIÓN ESTABAN TENIENDO UN IMPACTO GRAVE, CRÍTICO O MODERADO EN SU TRABAJO

La información errónea y la desinformación están afectando enormemente a la democracia, debilitando la confianza en las instituciones democráticas y en los medios de comunicación independientes, y mermando la participación en los asuntos políticos y públicos. Durante el ciclo electoral, la exposición a información falsa y engañosa puede privar a los votantes de la posibilidad de votar con conocimiento de causa. La difusión de información errónea y desinformación puede socavar la confianza pública en las instituciones electorales y en el propio proceso electoral —como el registro de votantes, los sondeos y los resultados— y puede provocar apatía entre los votantes o el rechazo de unos resultados electorales creíbles. Los Estados y los líderes políticos han demostrado ser poderosas fuentes de desinformación, al difundir falsedades de forma deliberada y estratégica para mantener o asegurar el poder, o socavar los procesos democráticos en otros países.³⁶

Los grupos marginados y vulnerables son también blanco frecuente de la información errónea y la desinformación y el discurso de odio, lo cual agrava su exclusión social, económica y política. Las mujeres candidatas, votantes, funcionarias electorales, periodistas y representantes de la sociedad civil son objeto de desinformación sexista en Internet.³⁷ Estos ataques socavan la participación política y debilitan las instituciones democráticas y los derechos humanos, incluida la libertad de expresión y de acceso a la información de estos grupos.³⁸ Esta debe ser una prioridad cada vez más urgente para la comunidad internacional, entre otras cosas porque más de 2.000 millones de votantes acudirán a las urnas en todo el mundo en 2024.

La información errónea y la desinformación también se cruzan entre las plataformas y los medios

de comunicación tradicionales y dentro de ellos, por lo que es más complejo aún darles seguimiento y resolverlas si no se detectan en el origen. La desinformación puede ser una táctica deliberada de los medios de comunicación con influencias ideológicas y cooptados por intereses políticos y empresariales.³⁹ Al mismo tiempo, el auge de las plataformas digitales ha precipitado un declive dramático de los medios de comunicación fiables e independientes. La audiencia de las noticias y los ingresos por publicidad han migrado en masa a las plataformas de Internet, una tendencia exacerbada por la pandemia de COVID-19. En algunas regiones o países se observa la “extinción de los medios de comunicación” o “desiertos de noticias”, donde las comunidades pierden fuentes locales fiables de noticias, lo que contribuye a contaminar el ecosistema informativo,⁴⁰ El “lavado de noticias”, que consiste en disfrazar los contenidos patrocinados para que parezcan noticias, a menudo no está bien señalado cuando se publica en plataformas digitales, lo que le confiere un barniz de legitimidad. Una vez recogida por otros medios de comunicación, citada por políticos o compartida ampliamente a través de plataformas, el origen de la información se vuelve cada vez más turbio y los consumidores de noticias no pueden distinguirla de los hechos objetivos.

La desinformación también está teniendo un impacto directo en el trabajo de las Naciones Unidas. Varios Coordinadores Residentes, enviados, mediadores y personal de mantenimiento de la paz han expresado su preocupación por el efecto que tiene la desinformación en la seguridad operacional, la eficacia y la capacidad de actuación de la Organización. En una encuesta realizada en 2022, el 70 % del personal de mantenimiento de la paz de las Naciones Unidas afirmó que la información errónea y la desinformación estaban teniendo un impacto grave, crítico o moderado en su trabajo, mientras que el 75 % afirmó que estaban afectando a su seguridad. La información errónea y la desinformación también pueden utilizarse para atacar a los actores humanitarios y obstaculizar las operaciones de salvamento en las zonas de conflicto.

FIGURA II

INTEGRIDAD DE LA INFORMACIÓN Y LOS OBJETIVOS DE DESARROLLO SOSTENIBLE

Como se muestra en la figura III, las amenazas a la integridad de la información pueden tener un impacto negativo en la consecución de los Objetivos de Desarrollo Sostenible.

<p>1 FIN DE LA POBREZA</p>	<p>La información errónea y la desinformación siguen teniendo consecuencias para los esfuerzos de erradicación de la pobreza y la economía mundial. Las dificultades económicas también pueden alimentar la difusión de detestables mentiras polarizadoras, por ejemplo sobre grupos marginados. La crisis del costo de la vida ha sido un terreno especialmente fértil para difundir desinformación en la que se culpa falsamente al cambio a las energías renovables de la subida de los costes energéticos o de la pérdida de puestos de trabajo, por ejemplo.</p>	<p>9 INDUSTRIA, INNOVACIÓN E INFRAESTRUCTURA</p>	<p>La información errónea y la desinformación y el discurso de odio, así como las respuestas excesivas a estos fenómenos, pueden acarrear un impacto perjudicial para la innovación, por ejemplo al limitar el potencial de los grupos marginados y hacer que los espacios digitales sean menos igualitarios e inclusivos.</p>
<p>2 HAMBRE CERO</p>	<p>Las amenazas a la integridad de la información pueden agravar el hambre en el mundo, entre otras cosas exacerbando los conflictos, el cambio climático, las catástrofes, la pobreza y la desigualdad. La desinformación puede desviar la atención y de los retos que plantean los conflictos para la seguridad alimentaria mundial.</p>	<p>10 REDUCCIÓN DE LAS DESIGUALDADES</p>	<p>La información errónea y la desinformación y el discurso de odio que se difunden en Internet polarizan a la sociedad y se ceban en las comunidades marginadas y vulnerables, lo cual puede provocar una mayor exclusión social, económica y política.</p>
<p>3 SALUD Y BIENESTAR</p>	<p>Durante la pandemia de enfermedad por coronavirus (COVID-19), una infodemia de información errónea y desinformación comprometió las medidas de salud pública y las campañas de vacunación. No ha desaparecido la amenaza para la salud y el bienestar de los niños derivada de la información errónea y la desinformación antivacunas.</p>	<p>11 CIUDADES Y COMUNIDADES SOSTENIBLES</p>	<p>Los esfuerzos por conseguir ciudades y comunidades más sostenibles pueden verse socavados por la desinformación que niega los efectos de la actividad humana en el medio ambiente o desvía la atención de ellos.</p>
<p>4 EDUCACIÓN DE CALIDAD</p>	<p>La información errónea y la desinformación y el discurso de odio pueden repercutir negativamente en el acceso a una educación de calidad, en particular para los grupos marginados, entre ellos las mujeres jóvenes y las niñas. Las campañas de acceso a la información y de alfabetización mediática digital para aumentar la resiliencia serán esenciales a la hora de limitar el impacto social de los peligros en línea.</p>	<p>12 PRODUCCIÓN Y CONSUMO RESPONSABLES</p>	<p>Los activistas que respaldan iniciativas para fomentar una economía circular e impulsar prácticas de cero desechos han sido blanco de campañas de discurso de odio y desinformación en Internet.</p>
<p>5 IGUALDAD DE GÉNERO</p>	<p>El discurso del odio y la desinformación basados en el género pretenden subyugar sistemáticamente a las mujeres silenciándolas y apartándolas de la esfera pública. Pueden tener consecuencias devastadoras, desde suprimir la voz de las mujeres y alimentar la autocensura, hasta causar daños profesionales y para la reputación e incitar a la violencia física.</p>	<p>13 ACCIÓN POR EL CLIMA</p>	<p>La desinformación climática, y la inercia que fomenta, merma los esfuerzos por tomar medidas urgentes para atajar la crisis climática, entre otras cosas impidiendo el cambio crucial que supone pasar del uso de combustibles fósiles contaminantes a las energías renovables y las inversiones urgentes en resiliencia climática.</p>
<p>6 AGUA LIMPIA Y SANEAMIENTO</p>	<p>Dos mil millones de personas viven sin servicios de abastecimiento de agua potable gestionados de manera segura. La información errónea y la desinformación sobre la salubridad del agua potable y la seguridad del saneamiento pueden tener consecuencias peligrosas para la salud^a.</p>	<p>14 VIDA SUBMARINA</p>	<p>La información errónea y la desinformación pueden repercutir negativamente en los esfuerzos de conservación y uso sostenible de los océanos, los mares y los recursos marinos.</p>
<p>7 ENERGÍA ASEQUIBLE Y NO CONTAMINANTE</p>	<p>La información errónea y la desinformación sobre el clima, en gran parte alimentadas por la industria de los combustibles fósiles, están perjudicando a la transición urgente hacia formas más limpias de producción de energía, dejándonos sin margen de maniobra para conseguir un futuro sostenible para todos.</p>	<p>15 VIDA DE ECOSISTEMAS TERRESTRES</p>	<p>Los activistas medioambientales que trabajan para proteger la vida de los ecosistemas terrestres han sido blanco de campañas de odio y desinformación en Internet, que han tenido consecuencias en la vida real. La información errónea y la desinformación sobre el clima están socavando los esfuerzos de acción climática^a.</p>
<p>8 TRABAJO DECENTE Y CRECIMIENTO ECONÓMICO</p>	<p>Hay estudios que mencionan los efectos perjudiciales que tienen la información errónea y la desinformación y la incitación al odio en la economía^b.</p>	<p>16 PAZ, JUSTICIA E INSTITUCIONES SÓLIDAS</p>	<p>La desinformación y el discurso de odio se han utilizado para influir en las elecciones y en la opinión pública y sembrar la confusión. Se han utilizado para debilitar a los adversarios, frustrar los esfuerzos de pacificación, incitar a la violencia, prolongar los conflictos y dañar la confianza en el estado de derecho. Los esfuerzos por promover sociedades pacíficas e inclusivas y el papel de las Naciones Unidas en apoyo de la paz y la seguridad se han visto seriamente afectados como consecuencia de ello^d.</p>
<p>17 ALIANZAS PARA LOGRAR LOS OBJETIVOS</p>	<p>Hay estudios que mencionan los efectos perjudiciales que tienen la información errónea y la desinformación y la incitación al odio en la economía^b.</p>	<p>17 ALIANZAS PARA LOGRAR LOS OBJETIVOS</p>	<p>La información errónea y la desinformación y el discurso de odio pueden poner trabas a las alianzas importantes para alcanzar los Objetivos, mientras que el desvío de recursos para solucionar el problema puede debilitar los esfuerzos para no dejar a nadie atrás.</p>

a Organización Mundial de la Salud y Fondo de las Naciones Unidas para la Infancia, *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs* (Ginebra, 2021).

b Véase Roberto Cavazos y CHEQ, "The economic cost of bad actors on the Internet: fake news, 2019"; y London Economics, "The cost of lies: assessing the human and financial impact of COVID-19 related online misinformation on the UK", December 2020.

c Global Witness, *Last Line of Defence: The Industries Causing the Climate Crisis and Attacks against Land and Environmental Defenders* (2021).

d A/77/288.

NÚMERO DE MESES TRANSCURRIDOS PARA SUMAR

100 MILLIONES

DE USUARIOS ACTIVOS
MENSUALES EN TODO EL MUNDO

Fuente: Similarweb, con datos de Sensor Tower



INSTAGRAM

30 MESES



TIKTOK

9 MESES



CHATGPT

2 MESES

¿Cómo reforzar la integridad de la información?

La información errónea y la desinformación y el discurso de odio no existen en el vacío. Se propagan cuando las personas se sienten excluidas e ignoradas, cuando sufren los efectos de la disparidad económica y cuando sufren un desencanto político. Las respuestas deben abordar estos retos del mundo real. Los esfuerzos por alcanzar los Objetivos de Desarrollo Sostenible son fundamentales para construir un mundo en el que pueda restablecerse la confianza.

Al elaborar las respuestas, es importante no perder de vista el enorme valor que aportan al mundo las plataformas digitales. Han revolucionado la comunicación de masas en tiempo real, permitiendo la difusión de información vital durante catástrofes naturales y pandemias. Han contribuido a movilizar el apoyo a los objetivos por los que luchan las Naciones Unidas, demostrando a menudo ser fuerzas positivas para la inclusión y la participación en la vida pública. Han conectado a comunidades geográficamente dispares de personas excluidas por otros motivos, como las que padecen enfermedades raras, y a un amplio abanico de activistas que trabajan para hacer del mundo un lugar mejor.

RESPUESTAS NORMATIVAS

La cuestión de si las plataformas digitales pueden y deben ser consideradas jurídicamente responsables del contenido que alojan ha sido objeto de un largo debate. En algunos contextos, las leyes existentes basadas en la difamación, el ciberacoso y el hostigamiento se han utilizado eficazmente para contrarrestar las amenazas a la integridad de la información sin imponer nuevas restricciones a la libertad de expresión (A/77/287, párr. 44). Además, recientemente se han desplegado ciertos esfuerzos legislativos para resolver el problema en el ámbito regional y nacional. Cabe destacar el marco legislativo adoptado por

la Unión Europea en 2022, que comprende la Ley de Servicios Digitales, la iniciativa sobre transparencia y segmentación de la publicidad política y el Código de Buenas Prácticas en materia de Desinformación. La Ley de Servicios Digitales establece nuevas normas para los usuarios, las plataformas digitales y las empresas que operan en línea en la Unión Europea. Las medidas apuntan a los contenidos, bienes y servicios ilegales en línea y ofrecen a los usuarios un mecanismo tanto para señalar el contenido ilegal como para impugnar las decisiones de los moderadores que vayan en su contra. Exigen que las plataformas digitales mejoren la transparencia, especialmente sobre el uso y la naturaleza de los algoritmos de recomendación, y que las plataformas más grandes faciliten a los investigadores el acceso a los datos.

El Código de Buenas Prácticas en materia de Desinformación establece principios y compromisos para que las plataformas en línea y el sector publicitario contrarresten la difusión de desinformación en línea en la Unión Europea, que sus signatarios acordaron aplicar.⁴¹ Entre ellos se incluyen compromisos voluntarios para contribuir a desmonetizar la desinformación, ya sea impidiendo la difusión de anuncios que contengan desinformación o evitando colocar anuncios junto a contenido que contenga desinformación. Los signatarios también acordaron identificar la publicidad política de forma más clara, con datos sobre el patrocinador, el gasto publicitario y el período de visualización, y crear bases de datos de anuncios políticos en las que se puedan realizar búsquedas. Además, se comprometieron a compartir información sobre comportamientos manipuladores y maliciosos utilizados para difundir desinformación (como cuentas falsas, uso de bots de amplificación, suplantación de identidad y ultrafalsificaciones maliciosas) detectados en sus plataformas y a actualizar y aplicar periódicamente políticas para atajarlos. Otros compromisos se centran en capacitar a los usuarios para reconocer, comprender y denunciar la desinformación y en

reforzar la colaboración con los verificadores de datos y facilitar a los investigadores un mejor acceso a los datos. La verdadera prueba de fuego de estos nuevos mecanismos será su aplicación.

Uno de los principales objetivos del Código de Buenas Prácticas en materia de Desinformación es mejorar la transparencia de las plataformas. En febrero de 2023, los signatarios del Código publicaron los primeros informes de referencia sobre el cumplimiento de los compromisos adquiridos. En ellos se informó sobre hasta qué punto se impidió que los ingresos publicitarios llegaran a los originarios de la desinformación y otros comportamientos manipuladores detectados, como el esfuerzo coordinado a gran escala para manipular a la opinión pública en varios países europeos sobre la guerra en Ucrania.⁴²

RESPUESTAS DE LAS PLATAFORMAS DIGITALES

Las plataformas digitales son muy diversas en cuanto a tamaño, función y estructura y han adoptado una amplia gama de respuestas para hacer frente a los daños. Varias de las grandes plataformas se han comprometido públicamente a respetar los Principios Rectores sobre las Empresas y los Derechos Humanos, pero siguen existiendo carencias en materia de política, transparencia y aplicación.⁴³ Algunas plataformas no hacen cumplir sus propias normas y, en diversos grados, permiten y amplifican las mentiras y el odio.⁴⁴ Los algoritmos creados para promover el modelo de lucro de las plataformas están diseñados para maximizar deliberadamente la participación y monopolizar la atención y tienden a llevar a los usuarios hacia contenidos polarizadores o provocadores.

La mayoría de las plataformas digitales cuentan con algún tipo de sistema de autorregulación, moderación o mecanismos de supervisión, pero la transparencia en torno a la política y la práctica de retirada de contenido sigue siendo complicada.⁴⁵ Las regiones y los idiomas invierten en estos mecanismos de forma muy desigual y los recursos se concentran principalmente en el Norte

Global, al igual que el cumplimiento de sus propias normas por parte de las plataformas. Según una encuesta reciente, la traducción de las herramientas de moderación y los mecanismos de supervisión a los idiomas locales es incompleta en todas las plataformas.⁴⁶ Al mismo tiempo, la tarea de moderación suele subcontratarse y lamentablemente carece de recursos en idiomas que no sean el inglés.⁴⁷ Los moderadores han mencionado cuestiones preocupantes relacionadas con los malos tratos, las normas laborales y los traumas secundarios.⁴⁸ Hablan de la continua exposición a contenido violento e inquietante y de que tienen pocos segundos para determinar si una publicación infringe la política de la empresa. Los sistemas automatizados de moderación de contenido pueden desempeñar un papel esencial, pero están expuestos a posibles sesgos basados en los datos y estructuras utilizados para entrenarlos. También presentan elevados porcentajes de error en inglés y porcentajes de acierto aún peores en otras lenguas. Varias plataformas digitales cuentan con equipos de confianza y seguridad, derechos humanos e integridad de la información, pero estos expertos no suelen estar incluidos en las primeras fases del desarrollo del producto y a menudo son los primeros puestos de trabajo que se suprimen en las medidas de ahorro.

ACCESO A LOS DATOS

El acceso de los investigadores a los datos también es una prioridad urgente a escala mundial. La investigación y los recursos existentes siguen estando muy sesgados hacia los Estados Unidos de América y hacia Europa. Con notables excepciones, como los informes sobre el mantenimiento de la paz de las Naciones Unidas en África y la misión internacional independiente de investigación sobre Myanmar,⁴⁹ y algunos reportajes de investigación y cobertura periodística,⁵⁰ se han publicado pocas investigaciones sobre el impacto en el resto del mundo. Esto se debe en parte a que los investigadores carecen de acceso a las plataformas y a los datos que hay en ellas. Las herramientas necesarias para una investigación eficaz de los limitados datos que facilitan las plataformas

también tienden a estar diseñadas pensando en la comercialización y la mayoría son prohibitivas. Si las plataformas pasaran de un planteamiento de "acceso previa solicitud" a otro de "divulgación por defecto", con las salvaguardias necesarias para la privacidad, los investigadores podrían evaluar adecuadamente los daños.

EMPODERAR AL USUARIO

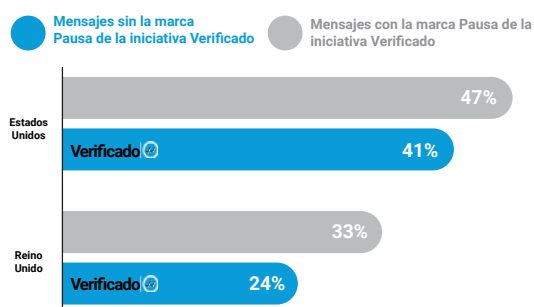
Varios grupos de la sociedad civil y académicos han llevado a cabo una amplia investigación sobre la mejor manera de abordar la información errónea y la desinformación y el discurso de odio, protegiendo al mismo tiempo la libertad de expresión. Algunos han subrayado la necesidad de encontrar soluciones de abajo arriba que empoderen a los usuarios de Internet para limitar el impacto de los daños en línea en la comunidad y descentralizar el poder de las manos de las plataformas.

Habría que alentar la participación de los usuarios, incluirlos e implicarlos en el espacio político, en particular a los grupos marginados. En concreto, los jóvenes poseen muchos conocimientos. Como nativos digitales, los jóvenes, en particular las mujeres jóvenes, y los niños ya son a menudo blanco de la información errónea y la desinformación y el discurso de odio y se verán directamente afectados por las nuevas plataformas. Los usuarios más jóvenes pueden hablar por experiencia

FIGURA III

CAMPAÑAS EFICACES DE LAS NACIONES UNIDAS CONTRA LA INFORMACIÓN ERRÓNEA Y LA DESINFORMACIÓN

Probabilidad de que alguien comparta noticias falsas (2021)^a



^a Según un estudio realizado por el Instituto Tecnológico de Massachusetts en marzo de 2021.

sobre el impacto diferenciado de las distintas propuestas y sus posibles defectos. También han contribuido activamente a los esfuerzos de promoción y de verificación de datos en línea.⁵¹

Una mejor capacidad de pensamiento crítico puede hacer a los usuarios más resistentes frente a la manipulación digital. En concreto, la alfabetización digital les enseña a evaluar mejor la información que encuentran en línea y a transmitirla de forma responsable. Diversas entidades de las Naciones Unidas han tenido valiosas experiencias en ese ámbito. La iniciativa Verified de las Naciones Unidas⁵² ha puesto en marcha una serie de tácticas, entre las que se incluyen mensajes dirigidos a los usuarios, advertencias a los usuarios sobre falsedades antes de que las encuentren (prebunking) y campañas de alfabetización digital.

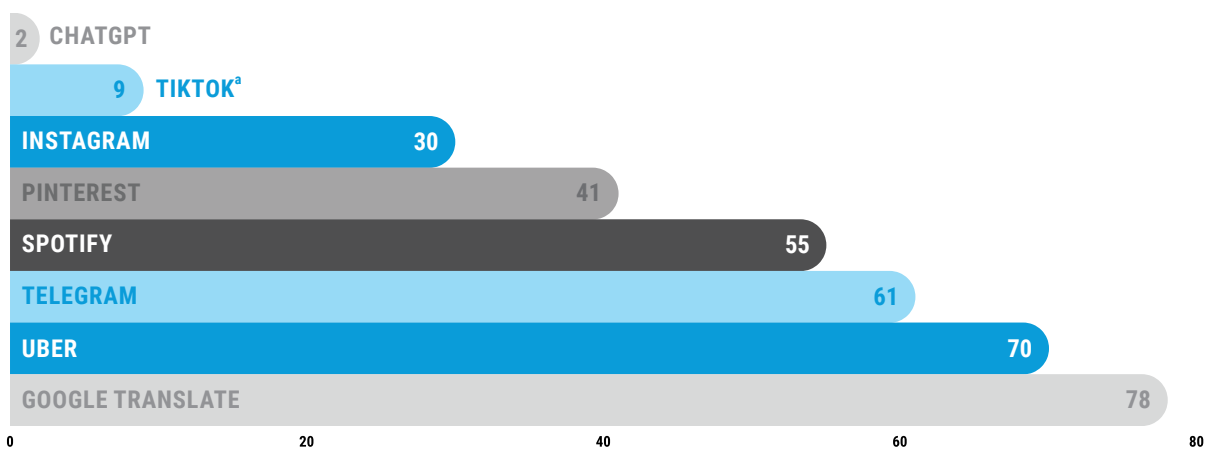
DESINCENTIVOS

Los modelos actuales de negocio de la mayoría de las plataformas digitales dan prioridad a la participación por encima de los derechos humanos, la privacidad y la seguridad. Esto incluye la monetización de los datos personales con fines lucrativos, a pesar de que cada vez hay más pruebas del daño social que causa este modelo de negocio. Algunos grupos de la sociedad civil e investigadores también han buscado vías para desmonetizar y, por tanto, desincentivar la creación y difusión de información errónea y la desinformación y discurso de odio en línea, señalando que si bien la libertad de expresión es un derecho humano fundamental, lucrarse con ella no lo es.⁵³ Hay propuestas que intentan abordar la rentabilidad de la desinformación, garantizar la plena transparencia en torno a la monetización del contenido y las evaluaciones de riesgo independientes y desincentivar a quienes participan en la publicidad en línea para que no permitan que haya desinformación.

Si una marca se anuncia junto a contenido de información errónea y desinformación y discurso de odio, se arriesga a que su campaña sea menos eficaz y, en última instancia, pone en peligro su reputación. Los anunciantes pueden desarrollar políticas claras para evitar financiar y legitimar inadvertidamente la información errónea y

FIGURA IV

NÚMERO DE MESES TRANSCURRIDOS PARA SUMAR 100 MILLONES DE USUARIOS ACTIVOS MENSUALES DE CHATGPT EN COMPARACIÓN CON OTRAS APLICACIONES CONOCIDAS



^a Meses para sumar 100 millones de usuarios activos mensuales incrementales después del lanzamiento global

Fuente: Similarweb, con datos de Sensor Tower.

desinformación y el discurso de odio y ayudar a que dejen de ser rentables. Las medidas de aplicación pueden incluir la gestión de listas actualizadas de inclusión y exclusión y el uso de herramientas de verificación de anuncios. Los anunciantes también pueden presionar a las plataformas digitales para que aumenten las medidas de protección de la integridad de la información y abstenerse de incluir anuncios en medios de comunicación que alimentan el odio y difunden desinformación.⁵⁴

MEDIOS INDEPENDIENTES

En decenas de países se han adoptado nuevas medidas que siguen socavando la libertad de prensa. Según el informe mundial de la UNESCO para 2022 de su emblemática serie Tendencias Mundiales en Libertad de Expresión y Desarrollo de los Medios de Comunicación, el 85 % de la población mundial experimentó un retroceso de la libertad de prensa en su país durante los cinco años anteriores.⁵⁵ Con 2.700 millones de personas aún desconectadas,⁵⁶ otra prioridad es reforzar los medios de comunicación independientes, impulsar la prevalencia de iniciativas de verificación de datos y apuntalar una información fiable y precisa en interés del público.

El verdadero debate público se basa en los hechos, relatados con claridad y presentados de manera ética e independiente. Los reporteros éticos, con una formación y unas condiciones de trabajo de calidad, pueden restablecer el equilibrio ante la información errónea y la desinformación. Pueden ofrecer un servicio vital: información precisa, objetiva y fiable sobre los temas que importan.

PREPARACIÓN PARA EL FUTURO

Aunque protejamos la integridad de la información en la actualidad, debemos asegurarnos de que las recomendaciones estén preparadas para el futuro, abordando las tecnologías emergentes y por venir. Lanzada en noviembre de 2022, la plataforma ChatGPT-3 de Open AI sumó 100 millones de usuarios en enero de 2023, convirtiéndose en la aplicación para consumidores de más rápido crecimiento de la historia,⁵⁷ muchas empresas se apuran desarrollando herramientas competidoras. Al tiempo que los recientes avances en inteligencia artificial —como los generadores de imágenes y la ultrafalsificación de videos— encierran un potencial casi inimaginable para enfrentar los retos mundiales, surge una preocupación grave y urgente por el potencial de amenaza a la

integridad de la información. Informes e investigaciones recientes demuestran que las herramientas de inteligencia artificial generativa crean información errónea y desinformación y discurso de odio, que se presentan de forma convincente a los usuarios como hechos.⁵⁸

Mi Enviado para la Tecnología dirige los esfuerzos para evaluar las implicaciones de la inteligencia artificial generativa y otras plataformas emergentes. Para ello debemos aprender de los errores del pasado. Las plataformas digitales se lanzaron al mundo sin suficiente conciencia ni evaluación del posible daño a las sociedades y las personas. Ahora tenemos la oportunidad de garantizar que la historia no se repita con la tecnología emergente. La era de la filosofía de Silicon Valley de “moverse rápido y romper cosas” debe llegar a su fin. Es esencial que la privacidad del usuario, la seguridad, la transparencia y la seguridad desde la fase del diseño se integren en todas las nuevas tecnologías y productos desde el principio.

RESPUESTAS DE LAS NACIONES UNIDAS

También se están tomando medidas, incluso por parte de las operaciones de paz de las Naciones Unidas y las oficinas en los países, para monitorear y analizar la amenaza de la información errónea y la desinformación para el cumplimiento de los mandatos de la Organización y para darle respuesta. La Estrategia y Plan de Acción de las Naciones Unidas para la Lucha contra el Discurso de Odio dispone la orientación estratégica con respecto al discurso de odio en los planos nacional y mundial. En febrero de 2023, la UNESCO celebró la conferencia Por un Internet Confiable, en la que se debatió un proyecto de directrices mundiales para regular las plataformas digitales, cuya finalización está prevista para finales de este año.⁵⁹

En conjunto, estas iniciativas y enfoques guían el camino para los principios subyacentes de un código de conducta de las Naciones Unidas.

MENCIONES MENSUALES DE

#CLIMATESCAM

EN TWITTER

100.000

NOVIEMBRE DE 2022

PICO DE LA COP27

200

FEBRERO DE 2022

MÁXIMO DE

199.300

PUBLICACIONES -
EN **ENERO DE 2023**

Hacia un código de conducta de las Naciones Unidas

El código de conducta de las Naciones Unidas para la integridad de la información en las plataformas digitales, que presentaré, se basaría en los siguientes principios:

- Compromiso con la integridad de la información
- Respeto de los derechos humanos
- Apoyo a los medios de comunicación independientes
- Mayor transparencia
- Empoderamiento del usuario
- Refuerzo de la investigación y del acceso a los datos
- Respuestas ampliadas
- Mayores desincentivos
- Mayor confianza y seguridad

Estos principios se han extraído de las ideas centrales debatidas en el presente informe de políticas y se ajustan y relacionan con el contenido de mi informe de políticas sobre el Pacto Digital Global. Se invitará a los Estados Miembros a aplicar el código de conducta a nivel nacional, cuyo contenido se perfeccionará a partir de las consultas con las partes interesadas, que también servirán para identificar metodologías concretas que lleven a la práctica sus principios.

El código de conducta puede basarse en las siguientes recomendaciones:

Compromiso con la integridad de la información

- a) Todas las partes interesadas deben abstenerse de utilizar, apoyar o amplificar la desinformación y el discurso de odio con

cualquier fin, incluido el de perseguir objetivos políticos, militares u otros objetivos estratégicos, incitar a la violencia, socavar los procesos democráticos o atacar a poblaciones civiles, grupos vulnerables, comunidades o personas;

Respeto de los derechos humanos

- b) Los Estados miembros deberían:
 - i) Garantizar que las respuestas a la información errónea y la desinformación y el discurso de odio se ajusten a las obligaciones que se derivan del derecho internacional, incluido el derecho internacional de los derechos humanos, y no se utilizan indebidamente para bloquear expresiones legítimas de puntos de vista u opiniones, incluso mediante apagones generales de Internet o prohibiciones dictadas contra plataformas o medios de comunicación;
 - ii) Adoptar medidas reguladoras para proteger los derechos fundamentales de los usuarios de plataformas digitales, incluidos mecanismos de aplicación, con total transparencia en cuanto a los requisitos impuestos a las empresas tecnológicas;
- c) Todas las partes interesadas deberían cumplir los Principios Rectores sobre las Empresas y los Derechos Humanos;

Apoyo a los medios de comunicación independientes

- d) Los Estados Miembros deberían garantizar un panorama mediático libre, viable, independiente y plural, con sólidas protecciones

para los periodistas y los medios de comunicación independientes, y apoyar la creación, financiación y formación de organizaciones independientes de verificación de datos en las lenguas locales;

- e) Los medios de comunicación deben garantizar una información independiente, precisa y ética, respaldada por una formación de calidad y unas condiciones de trabajo adecuadas, acordes con las normas y estándares internacionales de trabajo y derechos humanos;

Mayor transparencia

- f) Las plataformas digitales deberían:
 - i) Garantizar una transparencia significativa en relación con los algoritmos, los datos, la moderación de contenidos y la publicidad;
 - ii) Publicar y divulgar políticas accesibles sobre información errónea y la desinformación y el discurso de odio, e informar sobre la prevalencia de la desinformación coordinada en sus servicios y la eficacia de las políticas para contrarrestar tales operaciones;
- g) Los medios de comunicación deben garantizar una transparencia significativa de las fuentes de financiación y las políticas publicitarias y distinguir claramente el contenido editorial de la publicidad de pago, incluso cuando publiquen en plataformas digitales;

Empoderamiento del usuario

- h) Los Estados Miembros deberían garantizar el acceso público a una información gubernamental precisa, transparente y con fuentes creíbles, en particular la información que sirva al interés público, incluidos todos los aspectos de los Objetivos de Desarrollo Sostenible;
- i) Las plataformas digitales deberían garantizar la transparencia en el empoderamiento

y la protección de los usuarios, dándoles mayor capacidad de elección sobre los contenidos que ven y el uso que se hace de sus datos. Deberían permitir que los usuarios demuestren su identidad y autenticidad sin contrapartidas monetarias o de privacidad y establecer procesos transparentes de reclamación y denuncia respaldados por mecanismos de revisión de reclamaciones independientes, bien publicitados y accesibles;

- j) Todas las partes interesadas deberían invertir en programas sólidos de alfabetización digital para que los usuarios de todas las edades sepan cómo funcionan las plataformas digitales, cómo pueden utilizarse sus datos personales y cómo identificar y responder a la información errónea y la desinformación y el discurso de odio. Debe prestarse especial atención a garantizar que los jóvenes, los adolescentes y los niños conozcan perfectamente sus derechos en los espacios en línea;

Refuerzo de la investigación y del acceso a los datos

- k) Los Estados Miembros deberían invertir y apoyar la investigación independiente sobre la prevalencia y el impacto de la información errónea y la desinformación y el discurso de odio en todos los países y lenguas, especialmente en contextos desatendidos y en lenguas distintas del inglés, permitiendo a la sociedad civil y al mundo académico operar con libertad y seguridad;
- l) Las plataformas digitales deberían:
 - i) Permitir a investigadores y académicos acceder a los datos, respetando la privacidad de los usuarios. Debería permitirse a los investigadores recopilar ejemplos y datos cualitativos sobre personas y grupos víctimas de la información errónea y la desinformación y el discurso de odio para

comprender mejor el alcance y la naturaleza de los daños, respetando al mismo tiempo la protección de los datos y los derechos humanos;

ii) Garantizar la plena participación de la sociedad civil en la lucha contra la información errónea y la desinformación y el discurso de odio;

Respuestas ampliadas

m) Todas las partes interesadas deberían:

i) Asignar recursos para abordar los orígenes, la propagación y el impacto de la información errónea y la desinformación y el discurso de odio, respetando las normas de derechos humanos, e informar al respecto, e invertir más en capacidades de verificación de datos en todos los países y contextos

ii) Formar coaliciones amplias sobre la integridad de la información, que reúnan diferentes conocimientos y enfoques para salvar la distancia entre las organizaciones locales y las empresas tecnológicas que operan a escala mundial;

iii) Promover la formación y el desarrollo de la capacidad para comprender mejor cómo se manifiestan la información errónea y la desinformación y el discurso de odio y reforzar las estrategias de prevención y mitigación;

Mayores desincentivos

n) Las plataformas digitales deberían alejarse de los modelos de negocio que priorizan las interacciones por encima de los derechos humanos, la privacidad y la seguridad;

o) Los anunciantes y las plataformas digitales deberían garantizar que los anuncios no se coloquen junto a mensajes de información

errónea y desinformación y discurso de odio en línea, y que no se promueva la publicidad que contenga desinformación;

p) Los medios de comunicación deberían asegurarse de que toda la publicidad de pago y los contenidos publicitarios estén claramente marcados como tales y estén libres de información errónea y desinformación y discurso de odio;

Mayor confianza y seguridad

q) Las plataformas digitales deberían:

i) Garantizar la seguridad y la privacidad desde el diseño en todos los productos, entre otras cosas contratando a expertos en cuestiones relacionadas con la confianza y la seguridad, junto con una aplicación coherente de las políticas en todos los países e idiomas;

ii) Invertir en sistemas humanos y de inteligencia artificial de moderación de contenido en todos los idiomas utilizados en los países donde operan y garantizar que los mecanismos de denuncia de contenidos sean transparentes y que respondan rápidamente, especialmente en situaciones de conflicto;

r) Todas las partes interesadas deberían tomar medidas urgentes e inmediatas para garantizar un uso seguro, responsable, ético y respetuoso con los derechos humanos de la inteligencia artificial y abordar las implicaciones de los recientes avances en este campo en lo que respecta a la propagación de la información errónea y la desinformación y el discurso de odio.

Próximos pasos

- La Secretaría de las Naciones Unidas emprenderá amplias consultas con diversas partes interesadas sobre la elaboración del código de conducta de las Naciones Unidas, incluidos los mecanismos de seguimiento y aplicación. Esto podría incluir la creación de un observatorio independiente formado por expertos reconocidos para evaluar las medidas adoptadas por los actores que se comprometan con el código de conducta, y otros mecanismos de presentación de informes.
- Para apoyar y crear el código, la Secretaría de las Naciones Unidas quizás lleve a cabo estudios en profundidad para mejorar la comprensión de la integridad de la información a nivel mundial, especialmente en las partes del mundo poco investigadas.
- El Secretario General creará una capacidad específica en la Secretaría de las Naciones Unidas para ampliar la respuesta a la información errónea y la desinformación y el discurso de odio en línea que afecten a la ejecución del mandato y a las prioridades sustantivas de las Naciones Unidas. Basándose en el seguimiento y el análisis a cargo de expertos, esta capacidad desarrollaría estrategias de comunicación a medida para anticipar o abordar rápidamente las amenazas antes de que se conviertan en una espiral de daños en Internet y fuera de ella y apoyaría el desarrollo de capacidades del personal de las Naciones Unidas y de los Estados Miembros. Apoyaría los esfuerzos de los Estados Miembros, las plataformas digitales y otras partes interesadas para adherirse al código y aplicarlo cuando esté finalizado.

Conclusion

Reforzar la integridad de la información en las plataformas digitales es una prioridad urgente para la comunidad internacional. Desde la salud y la igualdad de género hasta la paz, la justicia, la educación y la acción climática, las medidas que limiten el impacto de la información errónea y la desinformación y el discurso de odio impulsarán los esfuerzos para lograr un futuro sostenible y no dejar a nadie atrás. Incluso actuando a escala nacional, estos problemas solo pueden resolverse totalmente con una mayor cooperación mundial. Las ideas centrales reseñadas en este documento demuestran que el camino

hacia una mayor integridad de la información debe basarse en los derechos humanos y ser multipartito y multidimensional. Se han condensado en una serie de principios que se tendrán en cuenta para un código de conducta de las Naciones Unidas para la integridad de la información en las plataformas digitales, que proporcionaría un modelo para reforzar la integridad de la información al tiempo que se defienden enérgicamente los derechos humanos. Aguando con interés la oportunidad de colaborar con los Estados Miembros y otras instancias para transformar estos principios en compromisos tangibles.

Anexo

CONSULTAS CON LOS ESTADOS MIEMBROS Y OTRAS PARTES INTERESADAS

Las ideas del presente informe se basan en las propuestas que se resumieron en el informe titulado "Nuestra Agenda Común" (A/75/982), y que se alimentaron de amplias consultas con los Estados Miembros, el sistema de las Naciones Unidas, líderes de opinión, jóvenes y miembros de la sociedad civil de todo el mundo. El informe político responde en particular a las ricas y detalladas reflexiones de los Estados Miembros y otras partes interesadas sobre "Nuestra Agenda Común", durante más de 25 debates de la Asamblea General.

Antes de la publicación del presente documento, se celebraron consultas con los Estados Miembros, entre otros, una reunión oficiosa con el Comité de Información, a la que se invitó a participar a todos los que no eran miembros del Comité. También se celebraron debates con asociados de la sociedad civil, académicos y expertos y con el sector privado, incluidas empresas tecnológicas.

Se llevarán a cabo amplias consultas para la elaboración del código de conducta antes de la Cumbre del Futuro.

Notas

- 1 [A/HRC/42/50; A/77/287; A/HRC/51/53](#); Naciones Unidas, "Statement by Alice Wairimu Nderitu, Special Adviser on the Prevention of Genocide, condemning the recent escalation of fighting in Ethiopia", [press release](#), comunicado de prensa, 19 de octubre de 2022; Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH), "Myanmar: Social media companies must stand up to junta's online terror campaign say UN experts", comunicado de prensa, 13 de marzo de 2023; ACNUDH, "Freedom of speech is not freedom to spread racial hatred on social media: UN experts", declaración, 6 de enero de 2023; Relatora Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, "#JournalistsToo: women journalists speak out", 24 de noviembre de 2021; y ACNUDH, "Sri Lanka: Experts dismayed by regressive steps, call for renewed UN scrutiny and efforts to ensure accountability", [press release](#), 5 de febrero de 2021.
- 2 [A/HRC/47/25](#), párr. 15.
- 3 Kalina Bontcheva y Julie Posetti, eds., *Un Acto de Equilibrio: La respuesta contra la desinformación digital y el respeto a la libertad de expresión - Informe de investigación de la Comisión sobre la Banda Ancha sobre "Libertad de expresión y lucha contra la desinformación en Internet"*, (Ginebra, Unión Internacional de Telecomunicaciones (UIT); París, UNESCO, 2020).
- 4 Véase Naciones Unidas, "Contrarrestar la desinformación", disponible en www.un.org/es/countering-disinformation, y [A/77/287](#).
- 5 Puede consultarse en www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf.
- 6 Las Naciones Unidas están llevando a cabo un estudio para examinar los vínculos y la relación entre la información errónea y la desinformación y el discurso de odio, y dónde convergen y divergen estos fenómenos relacionados pero distintos, tanto a nivel conceptual como operativo.
- 7 La Comisión Europea define las plataformas en línea en "Configurar el futuro digital de Europa: Plataformas en línea", 7 de junio de 2022. Disponible en <https://digital-strategy.ec.europa.eu/en/policies/online-platforms>.
- 8 Véase ACNUDH, "Moderating online content: fighting harm or silencing dissent?" 23 de julio de 2021., "[Moderating online content: fighting harm or silencing dissent?](#)", 23 July 2021.
- 9 Véase Naciones Unidas, "Contrarrestar la desinformación", y [A/77/287](#).
- 10 [A/HRC/47/25](#).
- 11 Stephanie Kirchgaessner y otros, "[Revealed: the hacking and disinformation team meddling in elections](#)", *The Guardian*, 14 de febrero de 2023.
- 12 Alexandre Alaphilippe y otros, "[Doppelganger – media clones serving Russian propaganda](#)", EU DisinfoLab, 27 de septiembre de 2022.
- 13 Red de Economistas de las Naciones Unidas, "[New economics for sustainable development: attention economy](#)".
- 14 Twitter, "[Acerca de Twitter Blue](#)"; y Meta, "[Testing Meta Verified to help creators establish their presence](#)", 17 de marzo de 2023.
- 15 [A/77/287](#).
- 16 En febrero de 2023, 173 Estados Miembros eran Estados partes en el Pacto Internacional de Derechos Civiles y Políticos.
- 17 Las limitaciones a la libertad de expresión deben cumplir las siguientes condiciones bien conocidas: legalidad, es decir, las restricciones deben estar previstas por ley de manera que se distinga entre expresión lícita e ilícita con suficiente precisión; necesidad y proporcionalidad, es decir, que la limitación imponga de forma demostrable la menor carga al ejercicio del derecho y proteja realmente, o pueda proteger, el interés legítimo del Estado en cuestión; y legitimidad, es decir, para ser lícitas, las restricciones deben proteger únicamente los intereses enumerados en el artículo 19 3) del Pacto Internacional de Derechos Civiles y Políticos
- 18 Por ejemplo, las comunicaciones audiovisuales en Côte d'Ivoire, Marruecos y Túnez, y el monitoreo de la incitación a la violencia del que se ocupa la Misión Multidimensional Integrada de Estabilización de las Naciones Unidas en la República Centroafricana.
- 19 Aleksii Knuutila, Lisa-Maria Neudert y Philip N. Howard, "[Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries](#)", *Harvard Kennedy School (HKS) Misinformation Review*, vol. 3, núm. 3 (abril de 2022).
- 20 UIT, *Medición de la Sociedad de la Información* (Ginebra, 2013).
- 21 Fondo de las Naciones Unidas para la Infancia (UNICEF), "[Protecting children online](#)", 23 de junio de 2022. Disponible en www.unicef.org/protection/violence-against-children-online.
- 22 UNESCO, documentos de trabajo sobre la gobernanza digital y los retos para la confianza y la seguridad. Disponible en www.unesco.org/en/internet-conference/working-papers.
- 23 Véase Julie Posetti y Kalina Bontcheva, "Disinfodemic: deciphering COVID-19 disinformation" [policy brief 1](#), (París, UNESCO, 2020), y "[Disinfodemic: dissecting responses to COVID-19 disinformation](#)" [policy brief 2](#), (París, UNESCO, 2020).
- 24 Center for Countering Digital Hate, *Pandemic Profiteers: The Business of Anti-Vaxx* (2021).
- 25 Michael A Gisondi y otros., "[A deadly infodemic: social media and the power of COVID-19 misinformation](#)", *Journal of Medical Internet Research*, vol. 24, núm. 2 (febrero de 2022).

- 26 [A/77/288](#).
- 27 [A/77/287](#), párr. 6.
- 28 En 2018, una misión internacional independiente de investigación designada por el Consejo de Derechos Humanos declaró que Facebook era "la principal plataforma de divulgación del discurso de odio en Myanmar" ([A/HRC/42/50](#), párr. 72).
- 29 Véase Noticias de las Naciones Unidas, "[Hate speech: a growing, international threat](#)", 28 de enero de 2023, y "[Digital technology, social media fuelling hate speech like never before, warns UN expert](#)", 20 de octubre de 2022
- 30 Véase John Cook, "Deconstructing climate science denial", en *Research Handbook in Communicating Climate Change*, David C. Holmes y Lucy M. Richardson, eds. (Cheltenham, Reino Unido, Edward Elgar, 2020). Cook informó de que Abraham et al. (2014) habían resumido cómo los artículos que contenían afirmaciones negacionistas, como afirmaciones de enfriamiento en las mediciones por satélite o estimaciones de baja sensibilidad climática, han sido refutadas sólidamente en publicaciones científicas. Del mismo modo, Benestad et al. (2016) intentaron replicar los hallazgos de trabajos contrarios y encontraron una serie de fallos, como el uso de métodos estadísticos inadecuados, dicotomías falsas y conclusiones basadas en conceptos físicos erróneos.
- 31 Global Witness, "[The climate divide: how Facebook's algorithm amplifies climate disinformation](#)", 28 de marzo de 2022.
- 32 Análisis del Departamento de Comunicación Global, con datos de Talkwalker.
- 33 Jeffrey A. Hicke y otros, "North America", en Grupo Intergubernamental de Expertos sobre el Cambio Climático, *Climate Change 2022: Impacts, Adaptation and Vulnerability*, Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge, Reino Unido, Cambridge University Press, 2022).
- 34 Mei Li, Gregory Trencher y Jusen Asuka, "[The clean energy claims of BP, Chevron, ExxonMobil and Shell: a mismatch between discourse, actions and investments](#)", *PLOS ONE*, publicación 17, núm. 2 (febrero de 2022).
- 35 Robert J. Brulle y Carter Werthman, "[The role of public relations firms in climate change politics](#)", *Climatic Change*, vol. 169, núm. 1-2 (noviembre de 2021). Según Global Disinformation Index, una organización sin ánimo de lucro dedicada a la vigilancia, en 2021 los anunciantes de la industria tecnológica proporcionaron 36,7 millones de dólares a 98 sitios web que incluían desinformación climática en inglés. Un informe de noviembre de 2022 del Center for Countering Digital Hate, un grupo de campañas, reveló que, solo en Google, casi la mitad de los 23,7 millones de dólares que gastaron en anuncios de búsqueda las empresas petroleras y de gas en los dos últimos años se destinaron a términos de búsqueda sobre sostenibilidad medioambiental. En 2020, InfluenceMap encontró 25.147 anuncios engañosos de 25 organizaciones del sector del petróleo y el gas en las plataformas de Facebook en Estados Unidos de América, con un gasto total de 9.597.376 dólares. Hasta ahora la respuesta ha sido desproporcionada a la magnitud del problema.
- 36 Véase la resolución [76/227](#); de la Asamblea General; la resolución [49/21](#) del Consejo de Derechos Humanos; y Acción Exterior de la Unión Europea "[Tackling disinformation, foreign information manipulation and interference](#)", 27 de octubre de 2021.
- 37 Lucina Di Meco, "[Monetizing misogyny: gendered disinformation and the undermining of women's rights and democracy globally](#)", #ShePersisted, febrero de 2023.
- 38 Véase Andrew Puddephatt, "[Social media and elections](#)", Cuadernos de Discusión de Comunicación e Información, No. 14 (Montevideo, UNESCO, 2019); y Julie Posetti y otras, "[The chilling: global trends in online violence against women journalists](#)", documento de debate e investigación (UNESCO, 2021).
- 39 EU Disinfo Lab, "[The role of 'media' in producing and spreading disinformation campaigns](#)", 13 de octubre de 2021.
- 40 Véase United Nations News, "[Social media poses 'existential threat' to traditional, trustworthy news: UNESCO](#)", 10 de marzo de 2022; y Anya Schiffrin y otras, "[Encontrar fondos para que el periodismo prospere: opciones de política para respaldar la viabilidad de los medios de comunicación](#)", Tendencias Mundiales en Libertad de Expresión y Desarrollo de los Medios (París, UNESCO, 2022).
- 41 Comisión Europea, "[Configurar el futuro digital de Europa: el Código de Buenas Prácticas en materia de Desinformación de 2022](#)", 4 de julio de 2022.
- 42 Véanse las observaciones de la Vicepresidenta de la Comisión Europea responsable de Valores y Transparencia, Věra Jourová, en Comisión Europea, "[Code of Practice on Disinformation: new Transparency Centre provides insights and data on online disinformation for the first time](#)", daily news, 9 February 2023. Disponible en https://ec.europa.eu/commission/presscorner/detail/en/mex_23_723.
- 43 Disponible en <https://unglobalcompact.org/library/2>.
- 44 Center for Countering Digital Hate and Human Rights Campaign, "[Digital hate: social media's role in amplifying dangerous lies about LGBTQ+ people](#)", 10 de agosto de 2022.
- 45 Véase Andrew Puddephatt, "[Letting the sun shine in: transparency and accountability in the digital age](#)", Tendencias mundiales en libertad de expresión y desarrollo de los medios (París, UNESCO, 2021).
- 46 Whose Knowledge?, Oxford Internet Institute y The Centre for Internet and Society, *Informe sobre el estado de las lenguas en Internet (2022)*.
- 47 [A/HRC/38/35](#).
- 48 Billy Perrigo, "[Inside Facebook's African sweatshop](#)", *Time*, 17 de febrero de 2022.
- 49 [A/HRC/42/50](#).
- 50 Algunos ejemplos notables son Maria Ressa, *How to Stand up to a Dictator* (Nueva York, HarperCollins, 2022); y Max Fischer, *The Chaos Machine* (Nueva York, Little, Brown and Company, 2022).
- 51 Véase UNICEF, "[Young reporters fact-checking COVID-19 information](#)".
- 52 Véase <https://shareverified.com/>.
- 53 Global Disinformation Index, un grupo sin ánimo de lucro, rastrea la publicidad que se coloca junto a la desinformación. Las Naciones Unidas han sido víctimas de esta práctica: Global Disinformation Index ha encontrado anuncios del UNICEF junto a artículos antivacunas y anuncios de la Oficina del Alto Comisionado de las Naciones Unidas para los Refugiados junto a contenido contra refugiados.
- 54 Conscious Advertising Network, manifiestos. Disponibles en www.consciousadnetwork.com/the-manifestos/.

- 55 UNESCO, *Journalism is a Public Good: World Trends in Freedom of Expression and Media Development – Global Report 2021/2022* (París, 2022).
- 56 UIT, "Facts and figures 2021: 2.9 billion people still offline", de noviembre de 2021. El **Pacto Digital Global** que los Estados Miembros adoptarán en la Cumbre del Futuro, que se celebrará en 2024, incluirá principios comunes que permitan lograr un futuro digital abierto, libre y seguro para todas las personas (véase www.un.org/techenvoy/global-digital-compact).
- 57 Krystal Hu, "ChatGPT sets record for fastest-growing user base – analyst note", Reuters, 2 de febrero de 2023.
- 58 Véase Center for Countering Digital Hate, "Misinformation on Bard, Google's new AI chat", 5 de abril de 2023; y Tiffany Hsu y Stuart A. Thompson, "Disinformation researchers raise alarms about A.I. chatbots", *The New York Times*, 13 de febrero de 2023.
- 59 El proyecto de directrices está disponible en www.unesco.org/en/internet-conference.

