

Informe sobre derechos humanos

Estadísticas y acciones
2022



Tabla de contenidos

1. Resumen ejecutivo	3	6. Novedades sobre la debida diligencia	34
2. Información sobre este informe de progreso	6	Filipinas	34
3. Prólogo	7	Israel y Palestina	35
4. Cómo Meta gestiona y rige el trabajo relativo a los derechos humanos	10	India	36
Capacitación para empleados de Meta	11	7. Participación de partes interesadas: cómo las partes externas fortalecen nuestro trabajo	38
Apelaciones y acceso a resarcimientos	11	Colaboración para generar espacios online seguros y respaldar la integridad de las elecciones en Nigeria	39
Consejo asesor de contenido	11	Participación de comunidades marginalizadas y subrepresentadas en la política de contenido	40
Global Network Initiative	12	Perspectivas de comunidades aborígenes sobre drogas sin fines medicinales	41
5. Cómo identificamos y abordamos riesgos significativos para los derechos humanos	13	Colaboración con organizaciones internacionales	41
Nuestros riesgos significativos	14	Ampliación de capacidades para identificar y escalar problemas con ayuda de socios de confianza	42
Libertad de opinión y expresión	15	Innovación por medio de foros de la comunidad	43
Privacidad	16	8. Idioma	44
Igualdad y no discriminación	19	9. Información destacada de la edición	46
Lenguaje que incita al odio	20	Protocolo de la política de crisis	46
Accesibilidad	20	Rusia y Ucrania	47
Vida, libertad y seguridad de la persona	22	Irán	48
Más estrategias para respetar el derecho a la vida, la libertad y la seguridad de la persona	23	10. Nuestro enfoque para los derechos humanos y el desarrollo responsable de productos	49
Protección para defensores de derechos humanos y periodistas	23	Metaverso	49
Explotación y trata de personas	23	Inteligencia artificial	50
Vigilancia por contratación	24	11. En adelante: Consideraciones para el futuro	53
Difusión de imágenes íntimas no consensuadas	24		
Protección de los intereses de los niños	25		
Derecho a la participación pública, al voto y a presentarse a una candidatura	26		
Más información sobre nuestro trabajo en elecciones nacionales	28		
Elecciones intermedias 2022 de los Estados Unidos	28		
Brasil	28		
Kenia	29		
Libertad de reunión y asociación	31		
Derecho a la salud	33		

1 Resumen ejecutivo



Este es el segundo informe anual sobre derechos humanos de Meta, que abarca el año calendario 2022. Aquí demostramos nuestro cumplimiento de los compromisos asumidos en nuestra [Política corporativa de derechos humanos](#), que se basa en los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#) (UNGP). Asimismo, detallamos cómo desarrollamos nuestro trabajo para respetar los derechos humanos, cómo evaluamos posibles repercusiones que podrían vincularse con nuestros servicios y qué hacemos para mitigar estos riesgos.

Con el transcurso de los años, las apps y los servicios de Meta contribuyeron al rápido crecimiento de movimientos populares, que lograron desafiar ortodoxias, respaldar derechos y, con ello, cambiar el mundo. El año 2022 no fue la excepción. A medida que surgen nuevos desafíos, nuestro trabajo en materia de derechos humanos evoluciona para abordar asuntos como el uso de nuestros servicios en países en conflicto, amenazas a la internet abierta y el auge de la inteligencia artificial.

Reforzamos nuestros sistemas de gobernanza para lograr mediante nuestra labor que se respeten los derechos humanos en todos nuestros productos y servicios. Esto incluye continuar concediendo autoridad al [Consejo asesor de contenido](#), que [publicó](#) 12 decisiones de moderación de contenido y realizó 91 recomendaciones en 2022. Nuestras iniciativas para reforzar la gobernanza conllevan también capacitar a nuestros empleados. Durante el año, lanzamos una capacitación sobre derechos humanos titulada *Bigger than Meta: Human Rights* en toda la empresa.

En 2022, Meta aprobó la [evaluación](#) de Global Network Initiative, que incluía un análisis de nuestras políticas y procesos, así como ocho casos de éxito a modo de ejemplo que implicaban solicitudes gubernamentales de eliminación y solicitudes de datos.

La debida diligencia respecto de los derechos humanos es fundamental para nuestra Política corporativa de derechos humanos. En este informe, incluimos hallazgos clave obtenidos a partir de nuestra [evaluación integral de riesgos significativos para los derechos humanos](#) (CSRA) y proporcionamos información reciente sobre nuestra implementación de las recomendaciones a partir de nuestra labor de debida diligencia.

En 2022, realizamos la CSRA en asociación con [Business for Social Responsibility](#) (BSR), una organización independiente con conocimientos en materia de derechos humanos. El análisis identificó los siguientes ocho riesgos potenciales significativos que se debían abordar con la más alta prioridad, conforme a los criterios de los UNGP: 1.) Libertad de opinión y expresión; 2.) Privacidad; 3.) Igualdad y no discriminación; 4.) Vida, libertad y seguridad de la persona; 5.) Protección de los intereses de los niños; 6.) Derecho a la participación pública, al voto y a presentarse a una candidatura; 7.) Libertad de reunión y asociación; y 8.) Derecho a la salud. La evaluación determinó que los riesgos potenciales en relación con la manera en que Meta brinda nuestros servicios y modelos de ingresos son variados y comprenden diversos matices.

A medida que surgen nuevos desafíos, nuestro trabajo en materia de derechos humanos evoluciona para abordar asuntos como el uso de nuestros servicios en países en conflicto, amenazas a la internet abierta y el auge de la inteligencia artificial.

Asimismo, describimos cómo aplicamos las recomendaciones obtenidas en procesos de debida diligencia sobre derechos humanos anteriores en relación con [Filipinas](#), [Israel y Palestina](#), y la India.

Además, seguimos abordando las recomendaciones detalladas en la [auditoría de derechos civiles](#) de 2021 de Meta. En 2022, logramos un progreso con 32 de las 117 recomendaciones e ítems pendientes, ya sea porque los implementamos o pusimos en marcha proyectos para trabajar en ellos, lo que significa que completamos un 84% de esta labor.

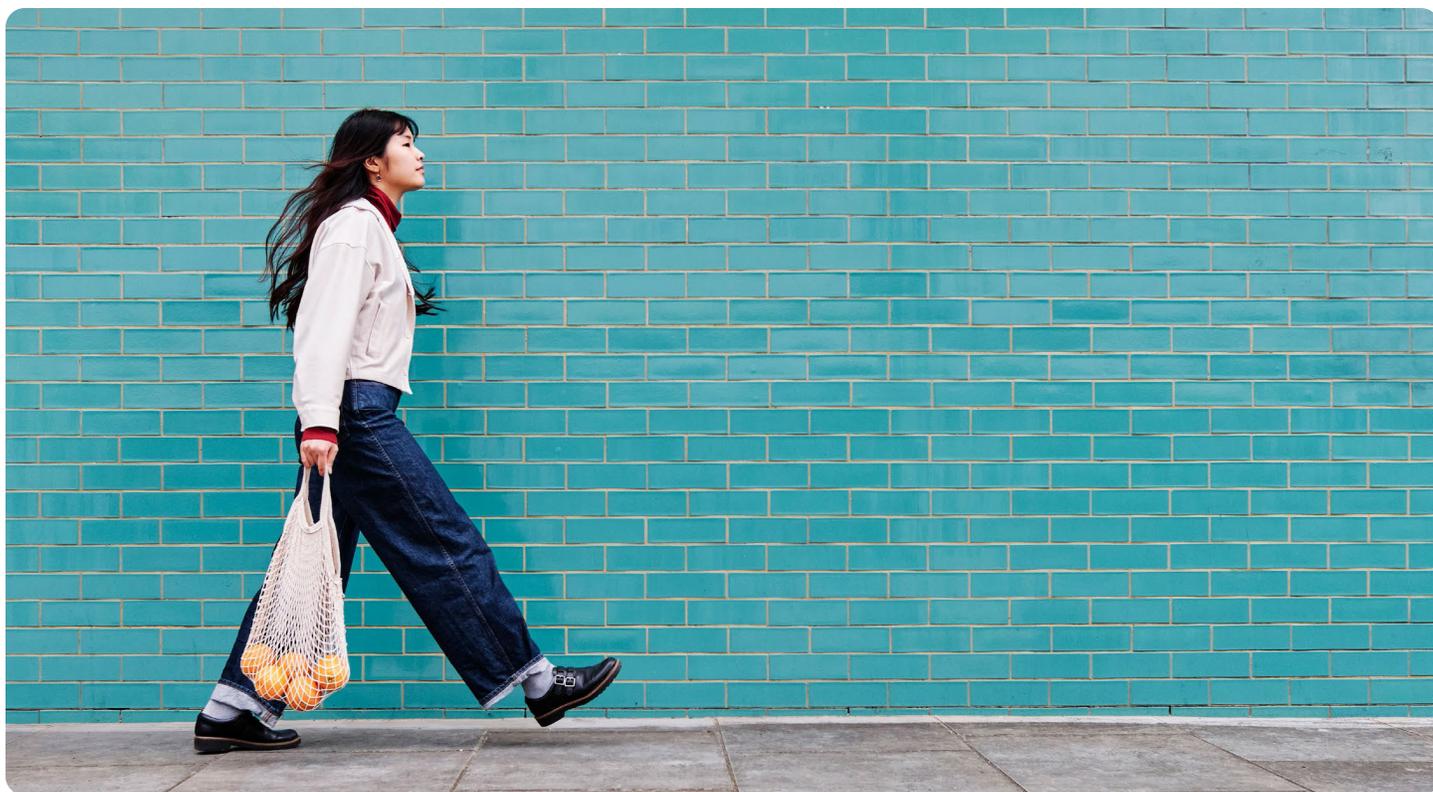
Escuchamos a las partes externas y aprendemos de ellas para cumplir con nuestras responsabilidades con los derechos humanos. La participación de partes interesadas forma parte intrínseca de nuestra Política corporativa de derechos humanos, y nos centramos en lograr una mayor participación de comunidades marginalizadas. El desarrollo de los servicios, las políticas de contenido y los procesos de moderación y las [Normas comunitarias](#) se basan en información valiosa obtenida de organizaciones de sociedad civil, [defensores de los derechos humanos](#), grupos marginalizados, organizaciones internacionales, [socios de confianza](#), inversionistas, anunciantes y usuarios.

Priorizamos forjar asociaciones en lugares en los que observamos los riesgos más significativos para los derechos humanos. Nuestra red de socios de confianza consta de más de 400 organizaciones no gubernamentales, sin fines de lucro, nacionales e internacionales ubicadas en 113 países que reportan contenido, cuentas y comportamientos que revisamos en contexto. En 2022, ampliamos nuestra red de socios de confianza a 36 nuevos países.

El contexto local es fundamental para entender los riesgos potenciales para los derechos humanos, y estamos avanzando para admitir más idiomas para poder lograrlo. Reconocemos que, durante momentos de crisis, los derechos humanos se encuentran bajo particular amenaza. Meta evalúa los riesgos de daño inminente, incluso antes de una crisis, y busca responder con políticas, servicios y acciones operativas específicas que respeten los derechos humanos. En este informe, revisamos el progreso de nuestro [Protocolo de la política de crisis](#) para garantizar que nuestra respuesta normativa a situaciones de crisis se base en principios y esté calibrada. Proporcionamos estadísticas de nuestra labor en Rusia, Ucrania e Irán, así como de las elecciones en Brasil, Kenia y Nigeria.

Seguimos trabajando para que los derechos humanos sean el eje de nuestro conocimiento e implementación de la tecnología de inteligencia artificial (IA), que avanza exponencialmente. Los modelos de IA pueden respaldar la capacidad de las personas de ejercer sus derechos, pero también pueden suponer sesgos problemáticos o impactos discriminatorios y generar contenido problemático. Nos comprometemos a abordar estos asuntos con una actitud de apertura y colaboración, y nos enorgullecen nuestras iniciativas de amplio alcance para resolver algunos de los problemas más complejos que conllevan los medios sociales y la IA.

Con pleno conocimiento de nuestras responsabilidades, nos esforzamos por implementar nuestros principios de derechos humanos a gran escala a favor de nuestros usuarios y la sociedad en general. Para lograrlo, brindamos a las personas el poder de crear comunidades.



2 Información sobre este informe de progreso

El segundo informe anual sobre derechos humanos de Meta¹ refleja el progreso de nuestros compromisos con los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#) y nuestra [Política corporativa de derechos humanos](#). Incluimos información que detalla cómo abordamos nuestro impacto en los derechos humanos. Esto incluye una revisión de nuestra reciente evaluación de riesgos significativos, las estadísticas obtenidas de nuestros procesos de debida diligencia en cuanto a derechos humanos, y las medidas que tomamos a modo de respuesta.



Nuestro [primer informe anual sobre derechos humanos](#) se publicó en julio de 2022 y cubría nuestros aprendizajes y el progreso alcanzado desde el 1 de enero de 2020 hasta el 31 de diciembre de 2021.

En este informe, el segundo que publicamos, incluimos nuestros aprendizajes y el progreso alcanzado desde el 1 de enero hasta el 31 de diciembre de 2022. Este toma como base principal el primer informe. Incluimos información sobre los servicios de Meta, como Facebook, Instagram, WhatsApp, Messenger y Reality Labs.

Lo complementamos con el [primer informe de prácticas comerciales responsables](#) de Meta, que concuerda con las normas de [Global Reporting Initiative](#) (GRI) y [Sustainable Accounting Standards Board](#) (SASB). Como se indica en el [informe de prácticas comerciales responsables de Meta de 2023](#), los derechos humanos se posicionan como un tema de altísima prioridad en la evaluación actualizada de temas prioritarios de Meta.

La Política corporativa de derechos humanos de Meta se aplica a toda la empresa. Cada servicio y entidad de Meta cuenta con políticas y procedimientos propios cuyo impacto en los derechos humanos varía por momentos. Este informe hace referencia a las medidas que toma Meta como empresa respecto de una o varias de las entidades de Meta. Las declaraciones no tienen como fin insinuar que Meta tomó esa misma medida respecto de todas las entidades.²

Durante la elaboración de este informe, nos vinculamos con muchas partes interesadas internas y externas, y buscamos equilibrar la brevedad con la integridad.



¹ El 28 de octubre de 2021, Facebook, Inc. cambió su nombre por Meta Platforms, Inc. Por consistencia, en este informe se utiliza el nombre "Meta" para referirse a la empresa tanto antes como después del 28 de octubre de 2021. Toda mención a "Facebook" se aplica solo a las apps y los servicios de medios sociales, no a la empresa en su totalidad.

² Por ejemplo, WhatsApp, una app de mensajes y llamadas con cifrado de extremo a extremo, tiene puntos de contacto de derechos humanos exclusivos. El debate sobre moderación de contenido y las medidas relacionadas en Facebook e Instagram que se detallan en este informe no se aplican a WhatsApp. Asimismo, a menos que se indique expresamente que una política o medida se aplica a WhatsApp, se considera que no se le aplica. Además, si bien muchas medidas descritas en este informe se aplican a Facebook e Instagram, se hacen distinciones intencionales en cuanto a las políticas y los procedimientos que recaen sobre estos servicios. Si una política se marca como política de "Facebook", es posible que no se aplique a Instagram. Ninguna declaración incluida en este informe tiene como objeto crear, ni debe interpretarse como que crea, nuevas obligaciones (legales o de otro tipo) respecto de la aplicación de una política o un procedimiento a otros servicios o entidades.



La misión de Meta es dar a la gente el poder de crear comunidades y unir más al mundo. Gracias a nuestras apps y servicios, los movimientos populares pueden prosperar y desafiar la ortodoxia y la autoridad establecida. Una manifestante en Irán, una activista LGBTQIA+ en Uganda, un periodista en Kiev: ellos, y todos los que usan nuestros servicios, son el motivo por el que nos esforzamos por respetar los derechos humanos.

Miles de millones de personas usan nuestras apps cada día, por lo que buscamos cumplir con las responsabilidades que conlleva operar a tan gran escala. Dada la velocidad con que acontecen los cambios tecnológicos y políticos, enfrentamos el desafío en constante evolución de entender y abordar cómo nuestros servicios afectan a los derechos humanos. Hace más de una década que Meta lidera en materia de investigación y desarrollo de IA, y adaptamos nuestro enfoque continuamente para cumplir con nuestras responsabilidades con los derechos humanos y prestar atención a riesgos emergentes, ya que poderosas herramientas como la IA generativa evolucionan con rapidez.

Los problemas de derechos humanos que enfrentamos son complejos, varían en función del país y la región, y se ven influenciados por la evolución de regulaciones

gubernamentales en todo el mundo. Por ejemplo, la [Ley de Servicios Digitales](#) de la Unión Europea, promulgada en 2022, genera para Meta importantes obligaciones de debida diligencia y elaboración de informes, pero se vincula explícitamente con el marco de derechos fundamentales de la región. Esto se desarrolla en el marco de la amenaza en evolución para la internet abierta y los derechos fundamentales de nuestros usuarios, que procede de la difusión de un modelo de internet autoritario, con ciudadanos segregados del resto de la internet global.

"Más de 3.500 millones de personas usan las apps de Meta —Facebook, Instagram, WhatsApp o Messenger— cada mes. Esto significa que las usa entre un tercio y la mitad de todos los seres humanos del planeta. Y cuando lo hacen, acceden a un universo de personas, ideas, noticias, comunidades y comercios interconectados, sin restricciones locales o nacionales. Nunca en la historia de la humanidad estuvimos tan conectados".

Nick Clegg, presidente de asuntos internacionales

En este informe, también compartimos el progreso logrado en cuanto a respetar los derechos humanos en todos nuestros programas, servicios y políticas. Seguimos explorando nuevas vías para incorporar los derechos humanos en nuestras prácticas y servicios. Esto incluye nuestra labor en cuanto a [fomentar la democracia](#), [asegurar un metaverso](#) abierto e interoperable, y encontrar nuevas maneras de garantizar que la [IA satisfaga las necesidades de las comunidades de un modo justo](#).

Todas nuestras iniciativas se basan en el compromiso que asumimos con nuestra [Política corporativa de derechos humanos](#) y los [Principios Rectores sobre las Empresas y los Derechos Humanos de la ONU](#) (UNGP). No llevamos adelante esta tarea solos. El [Consejo asesor de contenido](#) desempeña un papel fundamental y destacado en el sector como árbitro experto e independiente que revisa las decisiones de contenido y ofrece recomendaciones y opiniones de asesoramiento normativo. Conforme a su acta constitutiva, el Consejo asesor de contenido presta especial atención a las normas de derechos humanos que protegen la libertad de expresión.

Nuestra labor también se basa en los principios y las prácticas de organizaciones especializadas multilaterales y de la ONU, así como en alianzas con varias partes interesadas. Somos miembros del [Pacto Mundial de la ONU](#) y respaldamos sus principios. Asimismo, seguimos participando en el [Proyecto B-Tech](#) de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, que se centra en desarrollar recursos y pautas confiables para la implementación de los UNGP en el sector tecnológico. Como miembro de [Global Network Initiative](#)

(GNI), nos comprometemos con sus principios y pautas de implementación. En 2022, aprobamos la evaluación de GNI que se realiza tres veces al año.

También seguimos explorando nuevas formas de gobernanza. Por ejemplo, probamos los [foros de la comunidad](#) para reunir a grupos diversos de personas de todo el mundo con el fin de debatir problemas complejos y compartir perspectivas sobre un conjunto de recomendaciones.

Este informe tiene por objeto brindar mayor información sobre el trabajo que realizó Meta para identificar, mitigar y prevenir riesgos para los derechos humanos en 2022. Como siempre, nos queda mucho por hacer y esperamos poder brindar más informes de nuestro progreso de cara al futuro.

FIRMA:



Nick Clegg
Presidente de asuntos
internacionales



Jennifer Newstead
Directora ejecutiva de
asuntos legales

4 Cómo Meta gestiona y rige el trabajo relativo a los derechos humanos



Para avanzar en nuestra labor para respetar los derechos humanos en todos nuestros programas, servicios y políticas, debemos contar con estructuras de administración y gobernanza claras. Nuestro equipo de derechos humanos de Meta guía la implementación de la [Política corporativa de derechos humanos](#), que está bajo la supervisión del presidente de asuntos internacionales y la directora ejecutiva de asuntos legales.

Según se describe en nuestro [primer informe anual sobre derechos humanos](#), las tareas del equipo incluyen fomentar la integración de la política en políticas, programas y servicios actuales y en desarrollo; llevar a cabo procesos de debida diligencia; y respaldar la capacitación de los empleados en esta política. La política guía a los equipos para que desarrollen productos que respeten los derechos, respondan ante situaciones de emergencia y trabajen con rapidez y agilidad para incorporar los derechos humanos a gran escala.

Nuestra Política corporativa de derechos humanos nos compromete a presentar ante la junta directiva informes periódicos sobre asuntos clave relacionados con los derechos humanos. El comité asesor de riesgos y auditoría de la junta supervisa las diversas situaciones en la empresa que suponen una exposición a riesgos, incluidos riesgos relacionados con los derechos humanos, así como los pasos que la gerencia tomó para controlar o mitigar dicha exposición. El comité recibe información de manera periódica sobre la labor actual y en curso que realiza el equipo de derechos humanos.

Capacitación para empleados de Meta

En Meta, *cómo* creamos es tan importante como *qué* creamos. Nuestra capacitación sobre derechos humanos permite a los empleados entender mejor sus responsabilidades, así como adquirir los conocimientos y las competencias necesarios para cumplirlas.

Bigger than Meta: Human Rights, la capacitación que lanzamos en 2022, destaca el impacto potencial y real de los servicios, las políticas y las decisiones comerciales de Meta en cuanto a derechos humanos. Con ella buscamos fomentar una mentalidad orientada a los derechos humanos en nuestro trabajo cotidiano e instar a que estos se respeten en beneficio de todas las personas que usan nuestros servicios.

Esta capacitación complementa aquella sobre derechos civiles que Meta lanzó en julio de 2022, que se centra en los principios de justicia, igualdad y no discriminación.

Apelaciones y acceso a resarcimientos

En Meta, nos esforzamos por brindarles a las partes interesadas canales para reportar inquietudes, que Meta revisará y para las que proporcionará recursos y resarcimientos de conformidad con el UNGP n.º 31. Meta mantiene varios canales de reclamación, identificados en el [Código de conducta](#) y en el servicio de ayuda en las apps y plataformas, incluido un proceso de apelaciones ante el [Consejo asesor de contenido](#), primer organismo de este tipo.

Cuando se elimina contenido, notificamos a los usuarios e identificamos con claridad medios para [apelar](#) decisiones de eliminación de contenido. Los datos respecto de las decisiones de eliminación y otros aspectos de la aplicación de políticas se publican por medio del [Centro de transparencia](#). Para obtener más información, consulta nuestro [primer informe anual sobre derechos humanos](#).

Consejo asesor de contenido

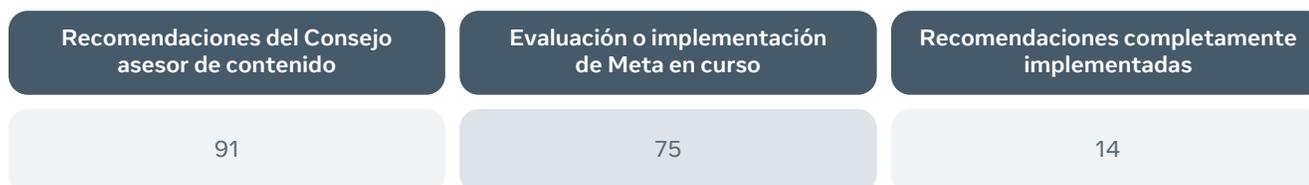
El Consejo asesor de contenido es una iniciativa única líder en el sector que busca [respetar el derecho a la libertad de expresión](#). Proporciona al público un mecanismo para presentar apelaciones adicionales ante un organismo externo. En 2022, Meta reforzó su compromiso con el Consejo asesor de contenido con la aprobación de una nueva subvención de tres años por \$150 millones al [fideicomiso del Consejo asesor de contenido](#). De este modo, el Consejo puede continuar trabajando como una entidad separada e independiente.

El Consejo asesor de contenido [toma decisiones de moderación de contenido y realiza recomendaciones](#) sobre servicios, operaciones y la política de contenido de manera absolutamente independiente. En 2022, el Consejo publicó 12 decisiones; nueve de ellas anulaban las decisiones de Meta. Estas abarcaban contenido en diversos países, incluidos Colombia, Etiopía, Afganistán, Croacia, Sudán, la India y los Estados Unidos. Los problemas abordados incluían, entre otros asuntos, representaciones de violencia gráfica en el contenido de un golpe militar, la mención del movimiento talibán en las

noticias y la descripción de violencia sexual contra menores de edad. Todas estas decisiones proporcionaron acceso a resarcimientos a los usuarios de la plataforma.

Además de sus decisiones, el Consejo asesor de contenido realizó 91 recomendaciones en 2022; 89 de ellas están pendientes de implementación, en proceso de implementación o completamente implementadas.

Recomendaciones del Consejo asesor de contenido en 2022



Global Network Initiative

Meta es miembro de [Global Network Initiative](#) (GNI) desde 2013. GNI es una creciente alianza entre empresas de internet y telecomunicaciones y diversos grupos de la sociedad civil, incluidos grupos de derechos humanos y libertad de prensa de todo el mundo. Nos comprometemos con los principios de libertad de expresión y privacidad de GNI ([Principios de GNI](#)), que se basan en leyes y normas de reconocimiento internacional. Estos compromisos se reafirmaron en nuestra Política corporativa de derechos humanos en 2021.

Todas las empresas que forman parte de GNI se someten a una [evaluación](#) independiente cada dos a tres años en la que se analiza su implementación de los principios.

En 2022, Meta aprobó la evaluación de GNI, que incluía un análisis de nuestras políticas y procesos, así como ocho casos de éxito a modo de ejemplo que implicaban solicitudes gubernamentales de eliminación y datos. El Consejo de GNI determinó de manera unánime que Meta había "hecho esfuerzos de buena fe para implementar los Principios de GNI, con una mejora a lo largo del tiempo", [estándar](#) con el que se evalúa a las empresas miembro. GNI emitirá un informe de evaluación público independiente que abarcará las evaluaciones de todas las empresas para el ciclo entre 2021 y 2022.

5 Cómo identificamos y abordamos riesgos significativos para los derechos humanos



En 2022, llevamos a cabo una evaluación integral de riesgos para los derechos humanos (CSRA), una herramienta básica del programa para guiar la gestión futura de riesgos. El propósito de la CSRA era identificar y priorizar los riesgos inherentes más significativos para los derechos humanos en la empresa en todo el mundo, que se definen como los impactos potencialmente adversos más importantes para los derechos humanos que genera Meta³ que afectan a usuarios y a otras personas sobre los que repercuten las acciones de Meta. Los riesgos "inherentes" son aquellos que existen antes de implementar procesos o medidas de mitigación, en contraposición a los riesgos "residuales", que son los que quedan luego de haber implementado procesos para abordar los riesgos inherentes.

Cómo se llevó a cabo: [Business for Social Responsibility \(BSR\)](#) realizó una investigación de fuentes secundarias y reuniones exhaustivas con partes interesadas internas y externas, lo que implicó más de una docena de reuniones con grupos de sociedad civil, representantes de la ONU e inversionistas de varias regiones y dimensiones de diversidad. La evaluación también recurrió a estadísticas obtenidas a partir de procesos de diligencia debida sobre derechos humanos y reuniones con partes interesadas anteriores. Esta priorizaba los riesgos en función de las definiciones de "gravedad" (alcance, escala, reparación) y "probabilidad", según se detallan en los UNGP y el marco de informes de los UNGP:

³ El término "impacto adverso para los derechos humanos" concuerda con lo que se define en los UNGP y se refiere al impacto que sufre una persona cuando una medida reduce o anula su capacidad de gozar de sus derechos humanos. Este informe no debe interpretarse como si toda persona particular o comunidad hubiera experimentado un impacto adverso, a menos que se indique específicamente.

1. **Alcance:** ¿Cuántas personas podrían verse afectadas por el impacto potencial adverso para los derechos humanos?
2. **Escala:** ¿Qué tan grave sería para las personas afectadas el impacto potencial adverso para los derechos humanos?
3. **Reparación/irreversibilidad:** ¿Una reparación regresaría a las personas afectadas a la posición en la que estaban o a una equivalente?
4. **Probabilidad de ocurrencia:** ¿Qué probabilidades hay de que ocurra el impacto potencial para los derechos humanos?

De conformidad con los UNGP, la CSRA prestó particular atención a los derechos, las necesidades y los desafíos de personas que podrían estar en mayor riesgo de ser vulneradas o marginalizadas.

Lo que aprendimos: el análisis identificó como prioridad más alta en las operaciones internacionales los siguientes riesgos significativos para los derechos humanos:

1. Libertad de opinión y expresión
2. Privacidad
3. Igualdad y no discriminación
4. Vida, libertad y seguridad de la persona
5. Protección de los intereses de los niños
6. Derecho a la participación pública, al voto y a presentarse a una candidatura
7. Libertad de reunión y asociación
8. Derecho a la salud

Nuestros riesgos significativos

La CSRA analizó los riesgos significativos inherentes teniendo en cuenta todos los derechos humanos de reconocimiento internacional en las operaciones de la empresa en todo el mundo y priorizó los ocho más significativos conforme a los criterios de los UNGP y al marco de informes de los UNGP. Estos riesgos se evalúan debajo, acompañados por ejemplos ilustrativos de algunos posibles factores de riesgo y una selección de las iniciativas de prevención y mitigación de Meta.



Libertad de opinión y expresión

El [derecho a la libertad de opinión y expresión](#) es fundamental para aquello en lo que creemos y trabajamos para proteger en Meta.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Interrupciones del servicio de internet y bloqueos en medios sociales que evitan que las personas ejerzan su libertad de expresión e impiden que intercambien noticias e información vitales

Límites excesivos sobre el contenido por parte del Gobierno

Aplicación de políticas de moderación de contenido de Meta que pueden limitar la libertad de expresión

Ejemplos de cómo Meta aborda el riesgo potencial

Recientemente [lanzamos](#) una función que permite a las personas conectarse con WhatsApp por medio de un servidor proxy cuando se interrumpe la conexión a internet o se bloquea el acceso a WhatsApp.

Evaluamos las solicitudes de eliminación gubernamentales válidas y legales conforme a los Principios de GNI, que se analizan de manera independiente en nuestra implementación con regularidad.

Aplicamos controles operativos exhaustivos para evaluar la validez de estas solicitudes y brindamos [información respecto de las que proceden del Gobierno en el Centro de transparencia](#).

Llevamos a cabo revisiones regulares de las políticas de contenido, con el aporte de diversas partes interesadas, para determinar si cumplen con las normas de derechos humanos.

Los modelos de inteligencia artificial predicen si un contenido es violento y gráfico o se considera lenguaje que incita al odio. Los sistemas independientes, como [nuestras tecnologías de aplicación de políticas](#), determinan si se debe tomar una medida, como eliminar o degradar el contenido, o remitirlo a un equipo de revisión manual para que lo analice en más detalle. Al aplicar nuestras políticas, pretendemos mitigar los riesgos de lenguaje que incita al odio, violencia e incitación y otros que pueden perjudicar los derechos humanos, pero sin socavar la libertad de expresión.



"Proteger en todo momento la privacidad es una inversión continua y colectiva en nuestra empresa, y todas las personas que somos parte de Meta tenemos la responsabilidad de impulsar nuestra misión".

Michel Protti, director ejecutivo de privacidad de productos

El [derecho a la privacidad](#) es fundamental para la misión de Meta y es condición necesaria para que se cumplan los demás derechos humanos, como la libertad de expresión, la libertad de reunión y asociación, y la libertad de culto y religión.

Nos comprometemos a mejorar y adaptar nuestras prácticas y políticas para anticiparnos a los desafíos relacionados con la privacidad de las personas en nuestras apps y servicios y abordarlos.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Solicitudes gubernamentales de datos de usuarios excesivas o innecesarias

Ejemplos de cómo Meta aborda el riesgo potencial

Protegemos a los usuarios de solicitudes gubernamentales excesivas o ilegales de datos por medio de nuestro equipo especializado de respuesta a las fuerzas del orden.

Rechazamos las solicitudes gubernamentales que no cumplen con normas reconocidas internacionalmente, incluidas aquellas estipuladas en nuestra [Política corporativa de derechos humanos](#), y las impugnamos ante un tribunal si es necesario.

Respondemos a solicitudes legales válidas en los casos en que creemos de buena fe que debemos hacerlo de acuerdo con la ley. También respondemos a determinadas solicitudes legales a las que no estamos obligados a responder por ley, pero creemos de buena fe que la respuesta:

- es una exigencia de la ley en otra jurisdicción en la que puede estar el usuario;
- afecta a los usuarios en esa jurisdicción; y
- es coherente con los estándares reconocidos a nivel internacional, incluida nuestra Política corporativa de derechos humanos.

(continuación)

(continuación)

Cuando cumplimos con [solicitudes gubernamentales de datos](#), nos esforzamos por notificar a las personas que se solicitan sus datos y publicamos un informe de transparencia sobre las solicitudes que recibimos del Gobierno.

Contenido o comportamientos en las apps de Meta que pueden afectar de manera adversa los derechos de privacidad y protección de datos (p. ej., doxing, extracción de datos)

Meta cuenta con sólidas Normas comunitarias en [Facebook](#) e [Instagram](#) que rigen cómo pueden usar nuestras apps los usuarios. No permitimos contenido que [comparte, ofrece o solicita información de identificación personal u otra información privada](#) que podría provocar daños físicos o financieros. Eliminamos contenido que infringe nuestras políticas tan pronto como tomamos conocimiento de este.

Utilizar la automatización para acceder a datos de las apps y los servicios de Meta o recopilarlos sin nuestro permiso ("[extracción de datos](#)") es una infracción de nuestras condiciones del servicio. Contamos con un equipo especializado, medidas técnicas y [procesos](#) que se centran en detectar, investigar y bloquear intentos de extracción de datos no autorizados.

Uso por parte de Meta o de un tercero de datos de usuario delicados o sumamente personales, incluida la información relacionada con características protegidas, con fines de publicidad segmentada de un modo que repercuta de manera negativa en los derechos de privacidad y protección

Toda la información que recopilamos y tratamos depende del servicio que estamos ofreciendo. Para que esta tarea nos resulte más sencilla, estamos desarrollando [tecnologías de mejora de la privacidad](#) que permiten minimizar la [recopilación y el uso de datos](#). Nuestro equipo se dedica a desarrollar estas tecnologías para que la utilicen diversos equipos de Meta, y estas se centran en áreas clave como eliminar información de identificación personal de los datos durante la recopilación y permitir a los equipos implementar el cifrado de extremo a extremo en nuestras apps y servicios.

En consulta con expertos en derechos civiles, legisladores y otras partes interesadas, implementamos cambios para que los usuarios tengan más control sobre los anuncios y [eliminamos opciones de segmentación de anuncios](#) relacionadas con temas que se pueden percibir como sensibles.

(continuación)

(continuación)

Colaboramos con el Departamento de Justicia de los Estados Unidos en el desarrollo y lanzamiento del sistema de reducción de varianza, una nueva tecnología que permitirá distribuir ciertos anuncios en nuestras apps de un modo más equitativo. Lanzamos este sistema en los Estados Unidos para los anuncios del sector inmobiliario y lo ampliaremos para abarcar anuncios de empleo y crédito.

La CSRA identificó varias vías de riesgo relacionadas con la publicidad segmentada, al tiempo que puso de manifiesto la naturaleza matizada de dichos riesgos. Por ejemplo, si bien los modelos comerciales basados en anuncios pueden poner en riesgo el derecho a la privacidad, es posible establecer mitigaciones para abordar dichos riesgos y, al mismo tiempo, preservar la amplia capacidad que tienen las personas de todo el mundo de expresar sus ideas y asociarse. Del mismo modo, la publicidad es un factor generador de oportunidades indispensable para pequeñas y medianas empresas que, de otro modo, carecerían de acceso viable a clientes potenciales.



Igualdad y no discriminación

El [derecho a la igualdad y la no discriminación](#) brinda igualdad de protección contra cualquier tipo de discriminación. Si bien la democratización de la tecnología sigue poniendo internet al alcance de una amplia variedad de personas, no todas las comunidades comparten las mismas experiencias online, en especial las personas y comunidades marginalizadas.

Nos esforzamos por crear un entorno online más inclusivo y equitativo para todos los usuarios por medio de nuestras Normas comunitarias, que prohíben el lenguaje que incita al odio, el bullying y el acoso. Nuestros equipos de derechos humanos y civiles trabajan en estrecha colaboración para procurar respetar la igualdad de derechos de las personas y mejorar la igualdad en todas nuestras tecnologías.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Contenido que afecta de manera adversa a la igualdad y la no discriminación (p. ej., lenguaje que incita al odio)

Ejemplos de cómo Meta aborda el riesgo potencial

Nuestras Normas comunitarias de Facebook, incluidas las políticas de [lenguaje que incita al odio](#), que se relacionan con características protegidas, definen lo que está y no está permitido en Facebook. Las [Normas comunitarias de Instagram](#) describen lo que está y no está permitido en Instagram. Tomamos medidas contra el contenido que infringe nuestras políticas. Nuestras [Normas de publicidad](#) brindan pautas y detalles normativos sobre los tipos de contenido que admitimos y los tipos de contenido del anuncio y segmentación que prohibimos.

Continuamos con [nuestra labor para desarrollar sistemas de IA responsable](#) a fin de reducir los sesgos y mejorar la igualdad, incluso en áreas como la entrega de anuncios, así como de disminuir la prevalencia de lenguaje dañino en apps de IA generativa. Dimos un importante paso para ayudar a los investigadores a medir la igualdad en nuestros modelos de IA. La falta de datos diversos, o datos que representen a una amplia variedad de personas y experiencias, puede conducir a que los resultados generados por la IA reflejen estereotipos problemáticos o no resulten equitativos para todos. Por ello, nos dedicamos a preparar [Casual Conversations v2](#),

(continuación)

(continuación)

un conjunto de datos basado en consentimiento y disponible públicamente que permite a los investigadores evaluar mejor la igualdad y solidez de determinados tipos de modelos de IA, con el objetivo de que sean más inclusivos.

La moderación en algunos idiomas y dialectos puede suponer un mayor desafío.

Usamos una combinación de IA y moderadores de contenido humanos para eliminar de nuestras apps contenido que infringe las políticas.

Utilizamos [traducciones manuales y automáticas](#) para mejorar la moderación del contenido. Estamos mejorando continuamente nuestros flujos de informes y clasificadores de integridad en los idiomas que admitimos.



Lenguaje que incita al odio

Eliminamos aproximadamente 50 millones de contenidos con lenguaje que incita al odio de Facebook en 2022; 91% de los cuales detectamos de forma proactiva y tomamos medidas al respecto. Eliminamos 16 millones de contenidos con lenguaje que incita al odio de Instagram; 92% de los cuales detectamos de forma proactiva y tomamos medidas al respecto. El lenguaje que incita al odio representaba menos del 0,02% del contenido visualizado en Facebook e Instagram.

Para obtener más información sobre cómo el equipo de derechos civiles de Meta aborda los asuntos de igualdad y no discriminación, consulta este [informe de progreso](#) y esta [actualización](#).

Accesibilidad

A medida que la tecnología avanza, es importante que las personas con discapacidades no queden rezagadas. Nos esforzamos por mejorar continuamente nuestras iniciativas a fin de garantizar la capacidad de acceso en nuestras apps y servicios. Que estos sean accesibles para personas con discapacidades es un compromiso clave de nuestra Política corporativa de derechos humanos, en la que citamos la [Convención de Derechos de las Personas con Discapacidad](#) como norma internacional fundamental.

Nuestras acciones: ampliar las iniciativas de todo el sector para mejorar la accesibilidad

Las tecnologías de realidad virtual, aumentada y mixta son particularmente prometedoras para las personas con discapacidades, ya sea porque ofrecen



de los 50 millones de contenidos con lenguaje que incita al odio eliminados de Facebook en 2022 se detectó de manera proactiva y se tomaron medidas al respecto

tecnologías de asistencia, o soluciones de rehabilitación o capacitación, o experiencias que impulsan el reconocimiento y la información. Meta sigue colaborando con [XR Association](#) en proyectos con el fin de mejorar el conocimiento en materia de accesibilidad y las iniciativas de distintos sectores, como el lanzamiento de un nuevo sitio centrado en accesibilidad, [xraccessibility.github.io](#), en colaboración con [XR Access](#). Las recientes innovaciones incluyen las gafas inteligentes [Ray-Ban Stories](#) con tecnología de Meta con las que es posible enviar mensajes y hacer llamadas en WhatsApp y Messenger sin usar las manos, y hacer capturas de imágenes y videos con comandos de voz y controles táctiles.

Además de incorporar funciones en nuestros servicios, Meta desarrolló y compartió recomendaciones técnicas para que los desarrolladores creen apps de realidad virtual accesibles. También trabajamos con la comunidad de personas con discapacidades y con organizaciones del sector en el desarrollo y lanzamiento de [pautas](#) de accesibilidad de realidad extendida para desarrolladores en asociación con XR Association.





Vida, libertad y seguridad de la persona

El [derecho a la vida, la libertad y la seguridad de la persona](#) se refiere a proteger la libertad contra el confinamiento físico, así como a proteger a las personas de lesiones físicas y psíquicas. Para Meta, respetar este derecho humano incluye abordar el riesgo de que el contenido provoque daños, como trata de personas, amenazas online amparadas por el Estado y la participación de grupos no pertenecientes al Estado en hechos de odio o violencia o el apoyo a estos.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Agentes malintencionados:

- Explotar las apps y los servicios de Meta para coordinar daños online y offline
- Usar indebidamente apps y servicios para realizar ciberataques o phishing
- Amenazar y acosar a [defensores de los derechos humanos](#), activistas y otros grupos vulnerables

Ejemplos de cómo Meta aborda el riesgo potencial

Nuestras políticas de seguridad se diseñaron para identificar amenazas adversas y proteger a las personas contra ellas. Meta cuenta con herramientas y recursos para periodistas y defensores de los derechos humanos, que se desarrollaron gracias al aporte directo de estos últimos.

Ampliamos la disponibilidad de [Facebook Protect](#) a más personas, en particular a defensores de los derechos humanos, para que tengan más control sobre sus cuentas, incluso a quién permiten ver sus publicaciones.

La [guía de seguridad](#) en nuestro [Centro de privacidad](#) centraliza una gama de recursos que ayudan a las personas a administrar su configuración de privacidad a fin de protegerlas de vigilancias online o accesos forzados o no autorizados. También les permite encontrar ayuda si experimentan bullying y acoso. Colaboramos con expertos y defensores de los derechos humanos en diversos asuntos para crear políticas, herramientas y recursos, que incluyan cultura digital, protección infantil y consejos para padres.

Meta invierte en reducir el uso indebido de nuestros servicios por parte de gobiernos totalitarios, grupos terroristas u otros agentes malintencionados que intentan detectar críticas al régimen y vigilar a figuras de la oposición y defensores de derechos humanos. Tomamos medidas si observamos dicha actividad, bloqueamos la infraestructura de su dominio para que no se comparta en nuestros servicios y, si es posible y corresponde, notificamos a las personas que creemos que fueron blanco de estas operaciones maliciosas.

Más maneras en que honramos el derecho a la vida, la libertad y la seguridad de la persona

Las siguientes iniciativas destacan nuestros enfoques para ayudar a grupos y personas en riesgo.

Protección para defensores de derechos humanos y periodistas

Desarrollamos herramientas y recursos en el centro de seguridad de Meta para defensores de derechos humanos, periodistas y otras personas vulnerables. Nuestra [guía de seguridad para periodistas](#) garantiza que estos estén equipados para gestionar su seguridad online con herramientas y recursos que protejan sus fuentes, contactos, inicios de sesión y su información personal. Asimismo, respaldamos a la comunidad periodística internacional patrocinando programas por medio de nuestro [Meta Journalism Project](#).

En 2022, tomamos estas medidas:

- Concedimos a activistas y periodistas más protecciones contra amenazas de violencia, lo que incluyó eliminar contenido que "deja en evidencia" que una persona es activista en situaciones que podrían ponerla en peligro.
- Protegimos las cuentas de defensores de derechos humanos que fueron arrestados o detenidos por gobiernos autoritarios o en entornos de conflicto para que fuerzas del orden locales, servicios de seguridad y agentes conflictivos no accedan a ellas sin autorización.
- Potenciamos [Human Rights Defender Fund y Journalist Safety Initiative de Meta en Asia-Pacífico](#), por medio de los cuales el grupo internacional de derechos humanos Civil Rights Defenders proporcionó evacuación urgente, reubicación temporal, medidas de seguridad digital, capacitaciones en seguridad y ayuda legal a más de 500 defensores de derechos humanos, principalmente en el sur y el sureste de Asia. Civil Rights Defenders también llevó a cabo evaluaciones de seguridad, facilitó la implementación de medidas de seguridad y proporcionó otro tipo de apoyo relacionado con la seguridad a defensores de derechos humanos y organizaciones en la región.
- Intensificamos la interacción con defensores de derechos humanos para aportar comentarios permanentes sobre el impacto de las amenazas que enfrentan estas personas tanto a nivel regional como mundial, y les brindamos capacitación sobre la guía de seguridad en el Centro de privacidad.
- Iniciamos un curso de capacitación gratuito online sobre protección y seguridad para periodistas y defensores de derechos humanos, creado y administrado por [International Center for Journalists](#) en asociación con Border Center for Journalists and Bloggers.

En 2022, presentamos la guía de seguridad en el Centro de privacidad ante la Oficina de la Relatora Especial de la ONU sobre la situación de los defensores de derechos humanos.

Explotación y trata de personas

En un esfuerzo por intervenir y evitar daños, eliminamos contenido que facilite o coordine la explotación de seres humanos, lo que incluye la trata de personas.⁴

En 2022, actualizamos nuestra [Política de explotación de personas](#) para prohibir el contenido que solicita servicios de tráfico ilegal. Asimismo, proporcionamos [recursos informativos](#) en Facebook sobre los riesgos de participar en tráfico ilegal, los indicios de posible explotación y vías para solicitar la migración legal, incluido el asilo. Desarrollamos estos recursos mediante consultas con expertos, incluida la [Organización Internacional para las Migraciones](#).

⁴ Para obtener más información sobre la explotación y trata de personas, consulta la página 32 de nuestro [primer informe anual sobre derechos humanos](#).

Seguimos permitiendo contenido que solicita o comparte información sobre seguridad personal y cruce de fronteras, y sobre cómo abandonar un país y solicitar asilo. Admitimos contenido que repudia o genera conciencia respecto de la trata de personas.



Líneas de ayuda en el estado de Bihar, India, para personas víctimas de trata

La trata de personas es un problema particularmente habitual en el estado indio de Bihar. Para asistir a las víctimas, Meta respaldó el [desarrollo de una línea de ayuda](#) en WhatsApp, que contribuye con el sistema de seguimiento del Departamento de Trabajo del estado de Bihar. Los usuarios pueden reportar incidentes de trabajo infantil a un número específico de WhatsApp, que brinda servicios en inglés, hindi y magahí. La línea de ayuda también se utiliza para supervisar el progreso de niños trabajadores rehabilitados en el estado.

Vigilancia por contratación

Con ayuda del sector de vigilancia por contratación, algunos gobiernos y agentes malintencionados tienen la capacidad para silenciar voces, ya que pueden dirigirse a activistas, disidentes, periodistas e investigadores e intimidarlos por medio de spyware abusivo. Este spyware pone como blanco a personas sin que estas lo sepan y sin respetar sus derechos humanos, libertades civiles y el estado de derecho.

Las prácticas de vigilancia por contratación abusan de nuestros servicios como un vector que implanta malware en los dispositivos móviles de nuestros usuarios. Logramos interrumpir y reportar muchas de estas operaciones. En 2022, colaboramos con socios del sector para proponer principios y prácticas de prevención y mitigación de esta amenaza. Lanzamos estos principios en la [Cumbre por la Democracia](#) bajo el auspicio de [Cybersecurity Tech Accord](#). En nuestras [recomendaciones normativas](#) de diciembre de 2022, ofrecimos un plan a nuestras partes interesadas, incluidos gobiernos, el sector de vigilancia por contratación y la sociedad civil, para unir fuerzas a fin de promover derechos y libertades online.

Difusión de imágenes íntimas no consensuadas

Compartir imágenes o videos íntimos de alguien, ya sea online u offline, sin su consentimiento es una infracción de sus derechos a la privacidad y la seguridad. Las imágenes se suelen difundir por varias apps y plataformas. Las imágenes íntimas no consensuadas se suelen denominar "porno de venganza", que se refiere a compartir contenido íntimo como un acto de "venganza" percibida contra una persona tras romper la relación.

Meta colabora con organizaciones sin fines de lucro y otras empresas para producir una herramienta gratuita, [StopNCII.org](#), que permite a posibles víctimas marcar imágenes para que las empresas que participan puedan detectarlas y eliminarlas y evitar así que se compartan online. Con una acción, las personas pueden solicitar que se eliminen imágenes íntimas no consensuadas de varias apps y plataformas. StopNCII.org, que se basa en tecnología desarrollada por Facebook e Instagram, está disponible en 22 idiomas y funciona en Facebook, Instagram y las plataformas de otras [empresas participantes](#), incluidas TikTok y Bumble.



Protección de los intereses de los niños

La [protección de los intereses de los niños](#) les concede a los menores el derecho a que se tengan en cuenta sus propios intereses en todas las acciones y decisiones que los atañen, lo que incluye protegerlos de la [trata de personas](#). El [marco](#) de protección de los intereses de los niños de Meta concuerda con los valores fundamentales de la [Convención sobre los Derechos del Niño de la ONU](#).

Queremos que los jóvenes disfruten de experiencias seguras y positivas online. [La protección de los niños online](#) es una de las principales prioridades de Meta. Asimismo, ofrecemos herramientas para que el tiempo que los adolescentes pasan online sea más significativo, incluidas más formas de centrarse y fijar límites, como recordatorios de descanso, consejos de expertos, modo silencioso, toques para cambiar de tema y la opción de ocultar la cantidad de "Me gusta" en las publicaciones del feed. El enfoque que usamos para desarrollar estas tecnologías se basa en investigaciones y comentarios directos de padres, adolescentes y expertos en salud mental y psicología infantil.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Niños expuestos a contenido no deseado o inapropiado

Ejemplos de cómo Meta aborda el riesgo potencial

Nuestras políticas exigen que toda persona que cree una cuenta en Facebook e Instagram tenga al menos 13 años. Redactamos nuestras políticas de un modo que nos permita ofrecer una experiencia segura a las personas mayores de 13 años, y consultamos con expertos en seguridad infantil mientras las desarrollamos. Restringimos cierto contenido en función de la edad (p. ej., determinados tipos de violencia gráfica); trabajamos para evitar recomendar determinados contenidos no infractores a menores (como aquellos que promocionan alcohol y tabaco); y tenemos un [control para contenido delicado](#) configurado con el ajuste predeterminado más estricto para menores. Prohibimos el contenido que sexualiza la infancia y, si tomamos conocimiento de una supuesta explotación sexual infantil, la reportamos ante el [Centro nacional de niños desaparecidos y explotados](#) (NCMEC) de los Estados Unidos.

También tenemos políticas de anuncios más restrictivas para los menores (p. ej., además de prohibir anuncios de bienes o servicios restringidos, como alcohol o tabaco, tampoco permitimos que se les muestren a los menores anuncios de productos para perder peso).

(continuación)

(continuación)

Nuestras políticas prohíben, y Facebook e Instagram buscan evitar, el contenido que fomente el suicidio, la autolesión, los trastornos alimentarios o el bullying y el acoso. Eliminamos dicho contenido apenas lo detectamos y seguimos mejorando la tecnología que se utiliza para identificarlo y eliminarlo.

Explotación sexual o abuso infantil

La sección de protección de los niños de nuestro [centro de seguridad](#) destaca el enfoque integral que adoptamos respecto de la seguridad de los niños, que incluye políticas que prohíben la explotación infantil, así como tecnología de vanguardia para prevenir, detectar, eliminar y reportar infracciones a las políticas.

Desarrollamos más de 30 herramientas para respaldar a niños, adolescentes y padres, incluidas, en Instagram, [herramientas de supervisión en nuestro Centro para familias](#) que permiten a padres y niños pasar tiempo online juntos.

Usamos tecnología de detección sofisticada en Facebook e Instagram con el fin de detectar, eliminar y prevenir la difusión de imágenes y videos de explotación infantil. Reportamos presuntas instancias de explotación infantil que aparecen en nuestros sitios ante el NCMEC y organismos de todo el mundo. En 2022, colaboramos con el NCMEC en el desarrollo de la [plataforma Take It Down](#) para menores a quienes les preocupa que sus imágenes íntimas se compartan en apps o plataformas online.



Derecho a la participación pública, al voto y a presentarse a una candidatura

El [derecho a la participación pública, al voto y a presentarse a una candidatura](#) en elecciones libres y equitativas es un pilar de la democracia. También es vital para el estado de derecho, la inclusión social, el desarrollo económico y el avance de los derechos humanos.

Proteger la integridad de las elecciones en nuestras apps y servicios es una de nuestras principales prioridades. Meta [invierte](#) una importante cantidad de esfuerzo y recursos para proteger las elecciones online antes del período electoral, durante este y una vez finalizado. Nuestro objetivo es brindar a las personas la capacidad de expresarse y participar en el proceso cívico, mientras luchamos contra el lenguaje que incita al odio, la injerencia y la influencia extranjera.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Contenido que repercute de manera negativa en la participación pública, el voto o la candidatura a un cargo público

Iniciativas para disuadir a las personas de votar, aumento de la cantidad de spam, injerencia extranjera o reportes de contenido que infringe nuestras políticas

Agentes malintencionados coordinados que interfieren en procesos electorales

Ejemplos de cómo Meta aborda el riesgo potencial

Se prohíben los intentos de interferir en el voto o suprimirlo. Trabajamos para identificar y eliminar este contenido, sin importar el lugar del que procede.

Proporcionamos un nivel de transparencia único en el sector en torno a publicidad y páginas relacionadas con la política, de modo que los usuarios puedan ver quién intentan influir en la votación. En algunos mercados, prohibimos los anuncios que disuaden a las personas de votar, claman de forma prematura la victoria en un proceso electoral, intentan deslegitimar la elección o no concuerdan con lo que las autoridades sanitarias consideran una votación segura.

El lenguaje que incita al odio afecta a los votantes. Nuestras [Normas comunitarias](#) abarcan políticas sobre acoso, incitación, violencia y lenguaje que incita al odio, y prohíben los ataques en función de características como el origen étnico o la religión. Eliminamos el contenido que infringe estas normas apenas lo detectamos.

Buscamos de forma activa contenido de organizaciones que Meta prohibió por intentar infringir nuestras políticas u ocasionar daños en la vida real.

En 2022, actualizamos [Facebook Protect](#), que ofrece herramientas de seguridad y protecciones adicionales a candidatos, sus campañas y funcionarios electos.

Para respaldar la seguridad de las elecciones, nuestras avanzadas operaciones de seguridad eliminan campañas manipulativas, identifican amenazas emergentes e implementan nuestra red de verificación de datos internacional líder del sector.

Usamos IA para identificar cuentas falsas, cuya amplia mayoría eliminamos antes de que se reporten. Eliminamos más de 150 redes de comportamiento no auténtico coordinado desde 2017.

Más información sobre nuestro trabajo en elecciones nacionales

En elecciones en las que hay mayor riesgo de contenido infractor, reunimos a un equipo especialista interfuncional con conocimientos en materia de integridad electoral, información errónea, seguridad, derechos humanos y ciberseguridad, y pericia lingüística en ese mercado. El equipo supervisa los riesgos emergentes en tiempo real y responde ante ellos. También controla las noticias y la actividad relacionada con las elecciones en otras redes sociales y medios tradicionales. Estas iniciativas nos aportan un panorama integral y nos permiten hacer un seguimiento del tipo de contenido que puede viralizarse.

En 2022, armamos y reforzamos equipos con el fin de estar preparados para las elecciones en distintas partes del mundo, como Brasil, Kenia y Nigeria, así como para las elecciones intermedias en los Estados Unidos. Realizamos inversiones, acudimos a partes interesadas y expertos, establecimos colaboraciones y mejoramos las tecnologías.



Elecciones intermedias 2022 de los Estados Unidos

En el período previo a estas elecciones, trabajamos en colaboración con funcionarios electorales estatales y locales para enviar alertas de voto en el feed, lo que generó más de 80 millones de notificaciones relacionadas con las elecciones. Combatimos operaciones nacionales y extranjeras, y expusimos e interrumpimos decenas de redes que intentaban interferir con el proceso electoral. Prohibimos más de mil movimientos sociales militarizados y eliminamos de nuestras apps cientos de miles de páginas, grupos y cuentas de QAnon. Asimismo, eliminamos el grupo Stop the Steal original y prohibimos más de 270 organizaciones que abogan por la supremacía blanca.

Brasil

Hubo elecciones generales en Brasil a comienzos de octubre de 2022 y, más tarde ese mismo mes, se llevó a cabo una segunda ronda de las elecciones presidenciales. Las elecciones se realizaron en medio de una crisis económica y de preocupaciones en torno a denuncias por ilegitimidad del proceso electoral. Nuestra labor en materia de integridad de las elecciones nunca cesa, e intensificamos nuestras iniciativas para las elecciones 2022 en Brasil un año antes. En preparación para las elecciones, nos centramos en cuatro áreas clave: prevenir la injerencia, colaborar con las autoridades electorales, luchar contra la información errónea y aumentar la transparencia en los anuncios.

Activamos nuestro centro de operaciones electorales para permitir que los especialistas en integridad de Brasil y de todo el mundo identifiquen posibles amenazas en nuestras apps y servicios en tiempo real. Desde agosto hasta octubre, aplicamos de manera proactiva procesos de revisión manual e IA para poder detectar y eliminar más de 310.000 contenidos de Facebook e Instagram de usuarios de Brasil que infringían nuestra [Política de violencia e incitación](#) y más de 290.000 contenidos que infringían la [Política de lenguaje que incita al odio](#). Además, redujimos la distribución

de contenido que nuestra tecnología identificó que podría infringir dichas políticas, a fin de evitar que se difunda con rapidez. Publicamos recordatorios del día de las elecciones en Facebook e Instagram, así como recordatorios para actualizar las tarjetas de identificación electoral.

Desde el 1 de enero hasta el 2 de octubre de 2022, más de 30 millones de personas hicieron clic en etiquetas de información electoral incluidas en publicaciones de Facebook sobre las elecciones en Brasil y se las redirigió a información oficial en el sitio web del Tribunal Superior Electoral (TSE). Más de 4,7 millones de personas se suscribieron al bot de chat del TSE en WhatsApp para recibir información fidedigna. Para combatir la violencia online contra las mujeres en la política, lanzamos una [guía](#) con el respaldo del TSE y la Women's Democracy Network en Brasil.

En la lucha con la información errónea, trabajamos (aún hoy) en colaboración con organismos de verificación de datos independientes a fin de comprobar la veracidad de publicaciones, reels, historias, videos, imágenes y comentarios en portugués en Facebook e Instagram. En 2022, nuestro programa de verificación de datos en Brasil pasó de tener cuatro socios a tener seis.

Para acabar con la información errónea, nuestros equipos internos detectaron y eliminaron redes que infringían nuestra Política de comportamiento no auténtico coordinado.

Para ampliar la transparencia de los anuncios, ampliamos las políticas de transparencia más allá de los anuncios políticos para abarcar también publicidad en Facebook e Instagram sobre temas sociales, como economía, seguridad y educación, en Brasil. Rechazamos 135.000 solicitudes de anuncios de fuentes nacionales y extranjeras en cuya segmentación se incluía también Brasil.

Consulta más información sobre las iniciativas que emprendimos durante las elecciones brasileñas [aquí](#) y [aquí](#).

Kenia

Las elecciones generales de Kenia se celebraron en agosto de 2022. Los procesos electorales anteriores se vieron entorpecidos por abusos a los derechos humanos. Como preparación para las elecciones de 2022, colaboramos con defensores de los derechos humanos clave, organismos de la ONU, periodistas y miembros de la sociedad civil para dar algunos pasos que nos permitieran reforzar nuestras medidas de seguridad e integridad, como ampliar nuestro programa de [socios de confianza](#).

Como ocurrió en elecciones similares, creamos un equipo interfuncional con conocimientos lingüísticos para que liderara nuestra labor en torno a las elecciones. Nuestras iniciativas se centraron en grupos vulnerables. Inscribimos a mujeres líderes en Facebook Protect para proteger cuentas de alto perfil, adoptamos medidas de seguridad más estrictas (incluida la supervisión para detectar



posibles amenazas de hackeo) y realizamos intervenciones para eliminar el acoso de Facebook e Instagram. Además, desarrollamos sistemas de detección y acción para proteger a periodistas y defensores de derechos humanos.

Organizamos una serie de sesiones de capacitación para defensores de derechos humanos y mujeres periodistas independientes. También trabajamos para combatir la difusión de información errónea, tanto dentro como fuera de la plataforma, eliminando contenido cuyo fin era suprimir el voto. Esta medida incluyó eliminar denuncias de hechos de violencia falsas o fuera de contexto. Basamos nuestras acciones, en parte, en el consejo de socios locales.

Incrementamos la inversión en nuestros verificadores de datos externos de Kenia, como AFP, Africa Check y PesaCheck. Asimismo, brindamos subvenciones a redacciones de noticias de Kenia para reforzar la verificación de datos en el idioma local. Para impulsar el conocimiento sobre lenguaje que incita al odio e información errónea, desarrollamos campañas en la plataforma y en radios locales en varios idiomas, y lanzamos dos servicios de asistencia, Trending Event y CrowdTangle Live Display, que permiten a verificadores de datos externos identificar posible información errónea y actuar al respecto.





Libertad de reunión y asociación

El [derecho a la libertad de reunión y asociación](#) es fundamental para la democracia y es la base de muchos otros derechos que garantizan las leyes internacionales, incluido el derecho a la libertad de expresión y a formar parte de asuntos públicos.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Contenido en las plataformas de Meta que puede generar en algunos usuarios la sensación de no poder reunirse libremente en las apps de Meta

Límites excesivos que imponen los gobiernos a contenido, que restringen el derecho de reunión y asociación

Ejemplos de cómo Meta aborda el riesgo potencial

El lenguaje que incita al odio, el bullying y otras formas de acoso pueden afectar el derecho de una persona a reunirse y asociarse. Eliminamos este tipo de contenido cuando la víctima lo reporta porque evita que las personas se sientan seguras y respetadas en las apps de Meta y en el mundo real. Meta cuenta con políticas exhaustivas contra el [bullying y el acoso](#), y la [violencia y la incitación](#), incluido contenido que fija como blanco a personas particulares por medio de páginas, grupos y eventos no deseados.

Buscamos eliminar el acoso dirigido a las masas porque afecta a varios derechos, incluido el derecho de reunión. Este incluye el acoso dirigido a personas que se encuentran en mayor riesgo de sufrir daños en la vida real, incluidos defensores de derechos humanos, menores, víctimas de tragedias o hechos violentos, y figuras de la oposición en países en riesgo durante períodos electorales.

La función de cifrado de extremo a extremo de WhatsApp protege el derecho a la privacidad y, por lo tanto, la capacidad de los usuarios de ejercer sus derechos de reunión y asociación.

Evaluamos las solicitudes de eliminación gubernamentales válidas y legales conforme a los Principios de Global Network Initiative, que se analizan de manera independiente en nuestra implementación con regularidad.

Aplicamos controles operativos exhaustivos para evaluar la validez de estas solicitudes y brindamos [información respecto de las que proceden del Gobierno en el Centro de transparencia](#).

(continuación)

(continuación)

Llevamos a cabo revisiones regulares de las políticas de contenido, con el aporte de diversas partes interesadas, para determinar si cumplen con las normas de derechos humanos.

Meta respeta el derecho de los empleados de nuestros proveedores a formar gremios. Si los empleados de un proveedor organizan un gremio, esto no afectará a nuestra decisión de trabajar o colaborar con dicho proveedor. Creemos que es importante que las empresas con las que trabajamos no se opongan al derecho de sus empleados de participar en gremios ni se lo impidan.





Derecho a la salud

Conforme a la [Organización Mundial de la Salud](#), "el goce del grado máximo de salud que se pueda lograr es uno de los derechos fundamentales de todo ser humano sin distinción de raza, religión, ideología política o condición económica o social".

Para honrar el [derecho a la salud](#), Meta aumenta el acceso a información de salud confiable, permite que personas con problemas de salud similares se conecten entre sí y les brinda herramientas para tomar decisiones fundamentadas sobre su salud y bienestar.

Ejemplos de posibles riesgos inherentes significativos para los derechos humanos identificados en la CSRA

Información errónea o desinformación en cuanto a la salud en las plataformas de Meta

Contenido que infringe las normas que incita a cometer daños en la vida real o se diseñó con ese fin

Ejemplos de cómo Meta aborda el riesgo potencial

En 2020, Meta adoptó una [política](#) por la cual elimina la desinformación respecto de la salud si:

- existe una emergencia de salud pública;
- las autoridades locales o mundiales líderes nos indican que una afirmación particular es falsa; y
- dichas organizaciones o autoridades nos informan que dicha afirmación puede contribuir de manera directa al riesgo de daño físico inminente.

Meta sigue usando herramientas de IA para que la labor de verificadores de datos tome mayor alcance, a fin de detectar información de salud falsa y engañosa, incluido mediante la clasificación de una imagen conforme a su contenido visual. Esto nos permite prohibir anuncios y publicaciones de comercio que promocionan determinados productos médicos, pero que infringen las normas.

En el centro de seguridad de Meta figuran nuestros [recursos de prevención de suicidios](#) y otra información útil para aquellas personas que puedan estar teniendo pensamientos suicidas.

Nuestras políticas prohíben, y Facebook e Instagram pretenden prevenir, el contenido que fomenta el suicidio, la autolesión, los trastornos alimenticios o el bullying y el acoso. Eliminamos dicho contenido tan pronto como lo detectamos y seguimos mejorando la tecnología que se utiliza para identificarlo y eliminarlo.

6 Novedades sobre la debida diligencia



Procuramos evaluar el impacto en los derechos humanos por medio de un proceso de debida diligencia respecto de los derechos humanos, según se establece en nuestra [Política corporativa de derechos humanos](#) y de conformidad con los [UNGP 17 y 21](#). Esto incluye tomar las medidas correspondientes e implementar los resultados, supervisar dicha implementación y elaborar informes anuales sobre estadísticas y medidas relevantes.

En 2022, publicamos los resultados obtenidos de evaluaciones de impacto en los derechos humanos y de ejercicios de debida diligencia respecto de los derechos humanos en [Israel y Palestina](#), la India y el [cifrado de extremo a extremo](#). Continuamos implementando las recomendaciones de estas y otras evaluaciones anteriores. También potenciamos nuestras iniciativas de debida diligencia para abordar el posible impacto del metaverso en los derechos humanos.

Filipinas

[Publicamos los resultados](#) de la evaluación de impacto en los derechos humanos independiente que encargamos en Filipinas en diciembre de 2021 y nuestra

respuesta a ella. Asimismo, en esta evaluación se realizaron recomendaciones respecto de temas como incitación a la violencia, vigilancia, explotación sexual online, tráfico, actividades extremistas y responsabilidad corporativa.

Debajo incluimos algunas de nuestras acciones posteriores:

- Creamos [nuevos servicios y desarrollamos políticas más estrictas](#) en colaboración con la Comisión electoral, organismos de control electoral, verificadores de datos independientes y organizaciones de sociedad civil durante las elecciones de 2022 en Filipinas. Gracias a estas iniciativas, logramos ampliar nuestra capacidad para eliminar contenido y redes, conectarnos con más personas mediante información confiable sobre las elecciones, fomentar la cultura digital y la participación cívica, y proporcionar más transparencia en los anuncios políticos.
- Pusimos la [biblioteca de anuncios de Facebook](#) a disposición en Filipinas y exigimos a los anunciantes que completen nuestro proceso de autorización de anuncios e incluyan el descargo de responsabilidad "Pagado por" en anuncios sobre elecciones, política y ciertas categorías de temas sociales.
- Implementamos varias [actualizaciones a nuestras Normas comunitarias](#), por ejemplo, brindar mayor [protección a figuras públicas](#) como periodistas y defensores de derechos humanos.
- Eliminamos más tipos de contenido infractor, y nuestras políticas ahora brindan [protecciones más firmes contra el acoso basado en género](#). Lanzamos nuevas políticas contra el [acoso masivo y la intimidación grupal](#), y eliminamos las iniciativas coordinadas de acoso masivo cuyo blanco son personas con mayor riesgo de sufrir daños en el mundo real. Esto incluye ataques contra disidentes.
- En colaboración con [International Center for Journalists](#) y [Border Center for Journalists and Bloggers](#), lanzamos un [programa de protección y seguridad digital gratuito](#) para ayudar a periodistas y defensores de derechos humanos a proteger sus activos digitales y contrarrestar el acoso online. El [centro de seguridad para periodistas](#) de Meta centraliza todos los recursos y las herramientas disponibles en nuestras apps.

Consulta más detalles al respecto en la [implementación de derechos humanos actualizada de 2023](#).

Israel y Palestina

Publicamos el informe de [Business for Social Responsibility](#) (BSR) procedente de su ejercicio de debida diligencia independiente en materia de derechos humanos sobre el impacto de nuestras políticas y procesos en Israel y Palestina durante mayo de 2021. Lo hicimos en respuesta a una solicitud del Consejo asesor de contenido en la que se nos pedía realizar una revisión independiente para determinar si la moderación de contenido de Meta en árabe y hebreo, incluido el uso de automatización, se había aplicado sin sesgos.

El informe reveló desafíos arraigados en todo el sector en cuanto a la moderación de contenido en regiones asediadas por conflictos, así como la necesidad de proteger la libertad de expresión y, al mismo tiempo, reducir el riesgo de que los servicios online se utilicen para transmitir odio o incitar a la violencia. Asimismo, puso de manifiesto la complejidad de gestionar problemas en torno al conflicto

debido a la dinámica social e histórica, la volatilidad de diversos acontecimientos violentos, y las acciones y actividades de organizaciones terroristas. BSR no obtuvo evidencia de sesgo intencional por parte de Meta en estos asuntos.

Durante el período del informe, tomamos algunas de estas medidas posteriores:

Política

- Revisamos nuestra definición del término "exaltar" en nuestra Política de personas y organizaciones peligrosas tras la exhaustiva colaboración con partes interesadas, como el sector académico, miembros de la sociedad civil y otros expertos en materia de terrorismo, derechos humanos y libertad de expresión de todo el mundo.

Aplicación de políticas

- Luego de una intensa evaluación interna, determinamos que crear una ruta especializada para una variedad de dialectos árabes en nuestros sistemas contribuiría a que nuestra moderación de contenido de gravedad alta en árabe fuese más precisa. Ahora estamos revisando el desarrollo de mecanismos para dirigir contenido por dialecto árabe de manera eficiente, a fin de mejorar la precisión de la revisión de contenido en árabe y prevenir mejor la aplicación insuficiente o excesiva de políticas.
- Llevamos a cabo un análisis para desarrollar un clasificador árabe específico para dialectos a fin de detectar contenido en ese idioma. En función de los resultados y el aporte de lingüistas y expertos en modelos de lenguaje, agregaremos una función de identificación de idiomas ampliada a nuestros sistemas que pueda reconocer contenido en diferentes dialectos árabes.
- Nos comprometemos a actualizar nuestros clasificadores para mejorar la precisión y el rendimiento con regularidad.
- Lanzamos un clasificador en hebreo que detecta de forma proactiva el contenido con lenguaje que incita al odio en este idioma y toma medidas al respecto.

Para obtener una actualización más detallada sobre el progreso logrado con la implementación de recomendaciones, consulta [Meta Update: Israel Palestine Due Diligence Update](#).

India

En el [primer informe anual sobre derechos humanos](#) de Meta, proporcionamos un resumen detallado de la evaluación de impacto en los derechos humanos de la India, que se desarrolló tras consultar las pautas de The Danish Institute for Human Rights sobre el análisis y los informes de la evaluación del impacto en los derechos humanos que tienen las actividades digitales.

En función de las pautas de este grupo y de otros expertos en derechos humanos, creemos que el método de divulgación mitiga los riesgos de seguridad conforme al UNGP n.º 21(c). En este informe, detallamos las medidas que ya tomamos o que tenemos previsto tomar. Seguiremos estudiando los resultados obtenidos y tendremos en cuenta sus recomendaciones como base para identificar y guiar medidas relacionadas. Haremos un seguimiento de nuestra implementación de forma regular a lo largo de 2023 y más allá.

Desde entonces, estos son los progresos conseguidos:

- **Participación de partes interesadas.** En respuesta a estadísticas relacionadas con el potencial de Meta para profundizar la participación de partes interesadas en la India de forma integral, Meta está definiendo mejor su enfoque respecto de la participación y el alcance de la sociedad civil, lo que incluye un programa específico de interacción con organizaciones comunitarias y de sociedad civil de dicho país.
- **Informes de transparencia.** Los informes que presentamos en la India ahora son más detallados. Además de nuestros informes sobre [solicitudes gubernamentales de datos de usuarios](#), seguimos proporcionando un [informe detallado en la India](#) cada mes, que incluye información sobre medidas tomadas contra contenido infractor en Facebook e Instagram, reclamaciones enviadas por los usuarios del país y órdenes que dicta el Grievance Appellate Committees conformado recientemente, cuya constitución se basa en las normas de TI. Seguimos perfeccionando nuestros sistemas para recopilar métricas sobre contenido eliminado a raíz de solicitudes gubernamentales e informes públicos; esto se refleja en los informes presentados ante el Consejo asesor de contenido en respuesta a sus recomendaciones ([cuarto trimestre de 2022](#) y [primer trimestre de 2023](#)).
- **Ampliación de colaboraciones.** Avanzamos mucho en nuestro compromiso para ampliar las colaboraciones con la sociedad civil de la India, incluido en materia de cultura digital, seguridad de mujeres y niños, y lucha contra el extremismo. Se relanzó el centro de seguridad para mujeres como [centro de seguridad de Meta](#) en 2023, que se tradujo en asamés, bengalí, gujarati, hindi, maratí, punyabí, tamil, telugu, canarés y malabar. Ampliamos la participación de organizaciones de la India en nuestra iniciativa de resiliencia y estamos preparando el lanzamiento de nuestro programa de redireccionamiento de búsqueda, que redirige a los usuarios que usan términos de búsqueda extremistas y relacionados con la violencia a recursos, información y grupos de intervención.
- **Moderación de contenido.** En respuesta a estadísticas relacionadas con políticas que buscan evitar que siga circulando contenido discriminatorio y de odio contra grupos vulnerables, Meta desarrolló una prueba de prototipos para la aplicación del Plan de Acción de Rabat al evaluar posible lenguaje que incita al odio. Adaptamos los principios conforme al Plan de Rabat en herramientas activas para la política de contenido, lo que incluye marcos basados en remisión para evaluar los conceptos de ataque (contra personas) y contenido que implica amenazas de uso de la fuerza por parte del Estado. Aumentamos la cantidad de grupos designados en todas las categorías de personas y organizaciones designadas en la India para asegurarnos de que personas malintencionadas no usen indebidamente nuestra plataforma.

Asimismo, Meta se [prepara](#) para las próximas elecciones estatales y generales tomando como base su experiencia con la labor de garantizar la integridad de las elecciones. Estas iniciativas incluyen activar nuestro centro de operaciones electorales, asegurar que los revisores de contenido brinden sus servicios en 20 idiomas indios, ampliar nuestra cartera de socios de verificación de datos independientes de 7 a 11 (que ahora abarcan 15 idiomas indios), exigir la transparencia de anuncios sobre política y trabajar en estrecha colaboración con autoridades electorales y de la sociedad civil.

7 Participación de partes interesadas: cómo las partes externas fortalecen nuestro trabajo



La interacción con partes externas de todo el mundo nos permite cumplir con nuestras obligaciones con los derechos humanos e impulsar con vigor la responsabilidad y la transparencia. En particular, reconocemos la importancia de mantener un diálogo significativo con partes interesadas de comunidades marginalizadas y subrepresentadas. La participación de partes interesadas es una pieza integral de nuestra [Política corporativa de derechos humanos](#) y de la labor que realizamos en este ámbito. Nos esforzamos por escuchar a expertos y activistas, miembros del sector académico y otras personas, y solicitamos su asesoramiento, y también los mantenemos al tanto de los desarrollos que Meta tiene en curso. Tomamos las estadísticas que aportan como base para el desarrollo y la aplicación de nuestras políticas de contenido, así como para nuestro enfoque respecto de las recomendaciones y la clasificación algorítmica.

En 2022, Meta reforzó y expandió nuestra colaboración permanente con partes externas con los siguientes grupos.

Partes interesadas	Iniciativas de consulta y colaboración
Organizaciones de sociedad civil	Respaldo en la revisión de las Normas comunitarias e iniciativas de integridad en áreas clave, como las elecciones
<u>Defensores de los derechos humanos</u>	Medidas para proteger la seguridad de la cuenta y minimizar la difusión de información errónea y contenido infractor
Grupos vulnerables o marginalizados	Mecanismos de comentarios que contribuyen al desarrollo de la política de contenido y la creación de servicios de seguridad
Organizaciones internacionales	Difusión de perspectivas sobre las políticas de derechos humanos y mejora de la comunicación en países de alta prioridad
<u>Socios de confianza</u>	Para este programa internacional, mejoras en lo que respecta a identificar y marcar contenido infractor, aporte de conocimientos locales a la aplicación de nuestras políticas, y desarrollo de servicios de seguridad
Inversionistas y anunciantes	Consultas e informes sobre la labor en materia de derechos humanos de Meta, incluida la diligencia debida en este ámbito
Usuarios	Aportes en asuntos controvertidos y toma de decisiones descentralizada

Debajo incluimos algunos ejemplos del trabajo realizado:

Colaboración para generar espacios online seguros y respaldar la integridad de las elecciones en Nigeria

En la etapa preliminar a las elecciones generales 2023 en Nigeria, colaboramos con una amplia cantidad de partes interesadas. Dichas colaboraciones se centraron en nuestras [políticas electorales](#) y las iniciativas para conservar la integridad y precisión de la información en nuestras apps. Recopilamos estadísticas sobre tendencias relacionadas con contenido infractor, lo que nos permitió mejorar nuestros mecanismos de detección y preservar la integridad de la información.

Nuestras acciones:

- Organizamos una capacitación en protección y seguridad para grupos vulnerables, incluidas mujeres que forman parte de la política o que son figuras públicas, activistas y defensores de derechos humanos.
- Llevamos a cabo una mesa redonda en Nigeria para tratar la [Política de personas y organizaciones peligrosas](#), así como dos mesas redondas en Londres con la diáspora nigeriana.
- Colaboramos con socios de confianza por medio de talleres y consultas para que estos puedan identificar y remitir contenido que incluye lenguaje que incita al odio, información errónea, violencia e incitación, injerencia en el voto, y bullying y acoso.

Participación de comunidades marginalizadas y subrepresentadas en la política de contenido

Progresamos en nuestro objetivo de incluir la voz de comunidades marginalizadas y subrepresentadas en el desarrollo de la política de contenido. Desarrollamos un [marco de inclusión](#) para garantizar que tengamos en cuenta a nuestras diversas partes interesadas en el desarrollo de políticas y que basemos las Normas comunitarias en su aporte.

Nuestras acciones:

- **Solicitamos comentarios sobre las Normas comunitarias.** Líderes de comunidades religiosas, aborígenes y étnicas con escasa representación en nuestras apps formaron parte de ocho mesas redondas realizadas en Europa, Medio Oriente y África para debatir qué tan efectivas son nuestras Normas comunitarias en cuanto a crear un entorno inclusivo en las apps y los servicios de Meta. Entre estas reuniones, se organizaron mesas redondas en Kenia después de las elecciones y se realizaron análisis de la crisis migratoria en Polonia.
- **Se fortaleció la relación con la comunidad LGBTQIA+.** Aumentamos la colaboración con grupos y representantes en Europa, Medio Oriente y África para abordar el impacto que las políticas de contenido de Meta tienen en los usuarios, en particular respecto del lenguaje que incita al odio y el bullying y el acoso. Entre ellos, había defensores de derechos humanos del colectivo LGBTQIA+, organizaciones de sociedad civil, miembros del sector académico y activistas del Reino Unido, España, Alemania, Kenia, Ghana, Namibia, Uganda, Egipto, Sudán, Zimbabue, Sudáfrica, Gambia y Tanzania.
- **Ampliamos nuestro alcance a la sociedad civil del Caribe.** Escuchamos las inquietudes respecto de la difusión de contenido potencialmente infractor y lideramos debates sobre derechos digitales, protección y seguridad online, seguridad de las mujeres y derechos de la comunidad LGBTQIA+. Expertos y grupos de sociedad civil de Haití, Trinidad y Tobago, Jamaica, Panamá, República Dominicana y las Islas Caimán participaron en la primera mesa redonda de este tipo.
- **Ampliamos la colaboración con comunidades religiosas.** Expandimos nuestras colaboraciones normativas para incluir a mujeres musulmanas en el Reino Unido, comunidades musulmanas en el Sáhel y la comunidad sij en el Reino Unido, así como la comunidad Bahá'í con representantes de Canadá, Estados Unidos, Chile, Brasil, Alemania, España, Suiza, Sudáfrica, Emiratos Árabes Unidos, Egipto y Túnez.

- **Elaboramos una guía sobre cómo interactuar con personas con discapacidades.**
En colaboración con organizaciones de derechos de las personas con discapacidades que trabajan en la convergencia de las discapacidades y la política de tecnología, elaboramos una guía para nuestros equipos de política de contenido.



Perspectivas de comunidades aborígenes sobre drogas sin fines medicinales

Colaboramos con partes interesadas en Norteamérica, Gabón, Camerún y Zimbabue para que contribuyan al desarrollo de la política sobre drogas sin fines medicinales en contextos religiosos o tradicionales. Exploramos cómo equilibrar la libertad de expresión respecto de sustancias tradicionales y religiosas con los riesgos de seguridad que implica la promoción potencial de drogas sin fines medicinales, como la ayahuasca. Entre las personas que nos asesoraron había una amplia variedad de miembros del sector académico y organizaciones de sociedad civil, incluidos especialistas que trabajan en el ámbito de salud, medicina, antropología, psicología clínica, política sobre drogas, etnografía, etnoornitología, religión e investigación de drogas psicodélicas, así como expertos legales, autoridades reguladoras, defensores de la libertad de expresión, profesionales de la salud, asociaciones de curanderos tradicionales y religiosos, líderes comunitarios y grupos aborígenes. Con las contribuciones recibidas, logramos revisar nuestro enfoque para poder mantener debates positivos sobre drogas sin fines medicinales en contextos tradicionales o religiosos.

Colaboración con organizaciones internacionales

El rápido crecimiento y el impacto de la tecnología instaron a muchas organizaciones internacionales a elaborar políticas y abordar cuestiones de derechos humanos en el mundo digital. Colaboramos de forma regular con una amplia variedad de organismos gubernamentales internacionales, incluidos la [Oficina del Alto Comisionado para los Derechos Humanos](#), la [UNESCO](#), la [Oficina para la Prevención del Genocidio y la Responsabilidad de Proteger](#), [UNICEF](#), el [Foro Económico Mundial](#), el [Foro de Gobernanza de Internet](#), la [ACNUR](#), la [OCDE](#), la [Fundación de las Naciones Unidas](#) y diversas organizaciones regionales. Proporcionamos información valiosa sobre nuestras Normas comunitarias, cómo se aplican y qué estrategias empleamos para proteger a los defensores de derechos humanos. En 2022, intensificamos nuestros compromisos con la ONU a nivel de país, incluido en Etiopía, Kenia, Ucrania, Haití y Afganistán.

Debajo incluimos algunos ejemplos de las acciones que realizamos:

- UNICEF y Meta trabajaron en estrecha colaboración para garantizar que los [ucranianos](#) tuviesen a disposición información crítica, junto con créditos publicitarios para costear mensajes de respaldo y recaudación de fondos en los principales mercados mundiales, Ucrania y países vecinos, como Polonia y Rumania.
- Nos reunimos de manera regular con funcionarios de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos para debatir asuntos mundiales y específicos de cada país, y participar de manera activa en el [Proyecto B-Tech](#).

- Proporcionamos capacitación en las Normas comunitarias y mecanismos de reporte a equipos de países de la ONU en Etiopía, Nigeria y Haití.
- Trabajamos en colaboración con la [Oficina del Enviado del Secretario General para la Tecnología](#) en lo respectivo al Pacto Digital Mundial y la necesidad de que este admita un enfoque internacional armonioso respecto de la gobernanza de internet que fomenta la libre circulación de datos. Esto es fundamental para proteger los derechos humanos online, incluidos los derechos a la privacidad y la libertad de expresión.
- Nos reunimos con funcionarios sénior de la ONU y diplomáticos en la Semana de Alto Nivel de la Asamblea General de la ONU para debatir cuestiones de derechos humanos y temas relacionados.
- Colaboramos con diversos Relatores Especiales de la ONU pertenecientes al [Consejo de Derechos Humanos de la ONU](#), incluidos aquellos cuya labor se centra en los defensores de derechos humanos, la libertad de expresión, Afganistán y Myanmar, entre otros, para compartir información y entender mejor sus inquietudes en lo que se refiere al rol que desempeñan los servicios de Meta.
- Seguimos respondiendo a las solicitudes de datos por parte de [mecanismos de investigación independientes de la ONU en Myanmar](#) que buscan respaldar las iniciativas de responsabilidad.

Ampliación de capacidades para identificar y remitir problemas con ayuda de socios de confianza

Priorizamos forjar asociaciones en regiones del mundo en las que observamos los desafíos más significativos para los derechos humanos.

Nuestra red de [socios de confianza](#) consta de más de 400 organizaciones no gubernamentales, sin fines de lucro, nacionales e internacionales ubicadas en 113 países que reportan contenido, cuentas y comportamientos que revisamos en contexto. Estos socios de confianza constituyen importantes aliados en nuestras iniciativas para entender el impacto local que tienen nuestros servicios. Nos permiten mantenernos al corriente de tendencias emergentes, pueden marcar contenido infractor que podemos pasar por alto y mejoran las políticas y su aplicación en ámbitos como daños e información errónea. Sus informes nos brindan valiosas estadísticas en las cuales basar el desarrollo de políticas y nos permiten proteger a nuestros usuarios.

Nuestras acciones:

- En 2022, ampliamos nuestra red de socios de confianza en 36 países, incluidos Yemen, Ucrania, Etiopía, Mali, Kenia, Nepal, Haití y Camboya.
- Nuestros socios nos alertaron sobre desarrollos y riesgos que, de otro modo, podríamos haber pasado por alto. Por ejemplo:
 - Los socios de confianza reportaron un aumento de sextorsión en nuestras apps en Yemen, lo que condujo a una investigación y, en última instancia, a la eliminación de una gran cantidad de usuarios de Facebook, cuentas comerciales y cuentas de Instagram.
 - En Camboya, estos socios marcaron publicaciones de Facebook en las que se reclutaban personas para trabajar en call centers en el Sudeste Asiático, que contenían indicadores clave de trata de personas y explotación laboral. Se eliminaron más de 30 publicaciones.
 - Nuestros socios de confianza en Pakistán marcaron información errónea sobre la Ley de Protección de Derechos de las Personas Transgénero y apuntaron contra contenido que ponía en riesgo a las personas. El contenido se eliminó, y tomamos medidas para asegurarnos de detectar otras instancias de contenido infractor similar en el futuro.

Innovación por medio de foros de la comunidad

Meta sigue explorando maneras de integrar más voces al proceso de toma de decisiones de Silicon Valley. Pusimos a prueba foros de la comunidad basados en democracia deliberativa, un estilo de toma de decisiones que emplean gobiernos y organizaciones de todo el mundo. Este postulado reúne a grupos representativos de personas para que entablen conversaciones estructuradas y tomen decisiones sobre asuntos polémicos o de particular complejidad.

Nuestras acciones:

En 2022, llevamos adelante un modelo de debate grupal basado en [Deliberative Polling®](#), un método creado en Stanford Deliberative Democracy Lab. Colaboramos con el laboratorio, así como con Behavioural Insights Team, un grupo de análisis de datos con sede en el Reino Unido, para garantizar que el proceso fuera justo e independiente.

Este foro de encuestas deliberativas fue el más importante de su tipo hasta la fecha gracias a la participación de más de 6.300 personas de 32 países. Su eje central es abordar el bullying y el acoso en espacios sociales virtuales.

Los participantes recibieron materiales educativos sobre el tema revisados de manera independiente, deliberaron entre ellos en varias sesiones de grupo reducido y tuvieron la oportunidad de plantear preguntas a expertos independientes de todo el mundo. Entre ellos había especialistas en ciberseguridad, abogados dedicados a los derechos humanos, activistas contra el bullying y el acoso, expertos en realidad virtual inmersiva y ex líderes de seguridad de la marca de Twitter, TikTok y Second Life, así como miembros del comité asesor de la ONU y del Foro Económico Mundial y ganadores de los premios a la libertad de prensa internacional.

Tras estas sesiones, se realizaron encuestas a los participantes para que votaran por las propuestas que habían debatido, a fin de basar en estas respuestas el desarrollo de experiencias virtuales en nuestras apps y servicios. Este evento fue parte de nuestra iniciativa general para crear programas que fomenten una toma de decisiones más inclusiva y permitan que más personas expresen su opinión en cuanto al desarrollo de nuestras apps y tecnologías.

En general, los participantes respaldaron con vehemencia la noción de que tanto creadores como propietarios de plataformas son responsables de supervisar los universos que crean, ya sean espacios públicos o exclusivos para miembros.

Los participantes apoyaron el uso de herramientas como detección automática de palabras, capturas de video y moderadores visibles e invisibles. Se hizo mayor hincapié en aplicar dichas herramientas en espacios públicos, más que en espacios exclusivos para miembros.

Para obtener información más detallada, consulta el [informe de resultados](#) de Stanford.



El idioma es uno de los factores más importantes que influyen en la repercusión de Meta en los derechos humanos. Se destacó la importancia del idioma en consultas con partes interesadas durante la [evaluación integral de riesgos para los derechos humanos](#). En 2022, aumentamos nuestra capacidad para admitir más idiomas y permitimos que más personas pudieran ejercer sus derechos online, incluido el [acceso a resarcimientos](#).

Para comprender el potencial de riesgos para los derechos humanos, es fundamental conocer el contexto local. Por tal motivo, la capacidad lingüística de Meta incluye traducción manual y automática. La compatibilidad de idiomas incluye detección de idiomas y clasificadores basados en idioma, moderación de contenido y traducción de interfaces de usuario, así como otro contenido, como el servicio de ayuda o las [Normas comunitarias](#).

Trabajamos continuamente para ampliar la compatibilidad de idiomas. Entre 2021 y finales de 2022, nuestro equipo incluyó otros 30 idiomas al servicio de ayuda profesional con la interfaz de usuario de Facebook. La lista incluye idiomas con poca representación en internet y cuyas fuentes lingüísticas y traducción representan un desafío, como javanés, shona, kirundi, cingalés, nepalí y asamés.

Asimismo, a finales de 2022, nuestras Normas comunitarias estaban disponibles en 77 idiomas.



No Language Left Behind

No Language Left Behind (NLLB) de Meta es un proyecto de IA revolucionario único en su tipo que utiliza modelos de código abierto capaces de ofrecer traducciones revisadas de alta calidad de manera directa entre 200 idiomas, incluidos aquellos menos comunes, como asturiano y luganda. El objetivo del proyecto es brindar a las personas la oportunidad de acceder a contenido web y compartirlo en su lengua materna, para poder comunicarse así con cualquier persona y en cualquier lugar, independientemente de su preferencia de idiomas.

La tecnología en la que se basa el modelo NLLB-200, ahora disponible mediante la herramienta de traducción de contenido de la Fundación Wikimedia, respalda a editores de Wikipedia con la traducción de información a su lengua materna e idiomas de preferencia. De este modo, se difunden más conocimientos en más idiomas a favor de los usuarios de Wikipedia de todo el mundo.



9 Información destacada de la edición



Durante momentos de crisis, los derechos humanos suelen encontrarse bajo particular amenaza. En épocas así, Meta evalúa los riesgos de daño inminente dentro y fuera de la plataforma, y responde con políticas, servicios y acciones operativas que respeten los derechos humanos.

En 2022, Meta fortaleció la respuesta de nuestra organización para respaldar a las personas durante períodos de tensión creciente en un país, incluido durante conflictos, elecciones o disturbios civiles. Para proporcionar esta respuesta, reunimos a un equipo interfuncional cuya labor es priorizar el apoyo y los recursos para anticiparnos y responder a situaciones de tensión con eficiencia en cualquier lugar del mundo. La labor se estructura en función de elecciones generales y países en los que se necesita una respuesta activa, así como también de la capacidad global de previsión de situaciones críticas futuras y la preparación para afrontarlas. Se desarrolló y perfeccionó un conjunto de normas de integridad que permiten a los usuarios reportar contenido infractor, que tiene en cuenta las competencias lingüísticas de cada país particular.

Protocolo de la política de crisis

En agosto de 2022, en función de la [recomendación](#) del Consejo asesor de contenido, lanzamos nuestro [Protocolo de la política de crisis](#) interno para garantizar que nuestra respuesta normativa ante situaciones de crisis se base en principios y esté calibrada.

El protocolo es un marco dinámico que nos permite identificar situaciones de crisis emergentes y evaluar su gravedad relativa. Guía nuestro uso ágil de medidas normativas excepcionales o específicas para mitigar el daño potencial de conformidad con los riesgos observados y en función de

intervenciones en situaciones de crisis anteriores. Como resultado, el protocolo nos ayuda a equilibrar una respuesta general consistente ante las crisis con la capacidad de adaptarnos con rapidez a las condiciones locales cambiantes. Para desarrollar el protocolo, consultamos con más de 50 expertos externos de todo el mundo versados en derechos humanos, prevención de conflictos, lenguaje que incita al odio, respuesta humanitaria y seguridad nacional. También llevamos a cabo una investigación original.

Este protocolo se diseñó para complementar e impulsar las iniciativas de respuesta ante crisis actuales de la empresa. El protocolo garantiza que Meta emplee un enfoque normativo integral e interdisciplinario para impulsar la mejora continua de nuestra respuesta ante crisis en los equipos de servicio, política y operaciones.

Entre nuestras actividades de respuesta ante crisis, citamos los siguientes ejemplos:

Rusia y Ucrania

La invasión rusa de Ucrania en febrero de 2022 supuso para Meta importantes desafíos en cuanto a [derechos humanos](#). En el ámbito digital proliferaban la propaganda y la desinformación con el fin de influir en la opinión pública, así como la vigilancia, las interrupciones de los servicios de internet y la promoción de contenido infractor. Necesitábamos acelerar nuestras iniciativas de moderación de contenido y, al mismo tiempo, equilibrar la libertad de expresión a fin de ayudar a la población ucraniana a coordinar iniciativas de ayuda humanitaria y expresar su resistencia a las fuerzas invasoras.

En Rusia, Facebook e Instagram quedaron bloqueados, y las autoridades incluyeron a Meta Platforms, Inc. en la lista de organizaciones extremistas. No obstante, WhatsApp continuó disponible.

Diligencia debida:

Desde el comienzo de la guerra, nuestra labor de debida diligencia en cuanto a derechos humanos incluyó interpretar leyes humanitarias internacionales pertinentes y normas de derechos humanos para que nuestros equipos pudieran transitar una situación de tan complejas características. Estas iniciativas incluyeron revisar videos, identificar prisioneros de guerra, así como aquellas personas que fomentan el lenguaje que incita al odio, la desinformación y la propaganda controlada por el Estado ruso con la que niegan la invasión a Ucrania.

Colaboramos con [defensores de derechos humanos](#), activistas, periodistas y grupos de sociedad civil y derechos humanos tanto en Rusia como en Ucrania para entender los problemas que enfrentaban. También forjamos un vínculo con la [ONU y otras instituciones internacionales](#) para profundizar nuestro conocimiento sobre los asuntos en cuestión y obtener más información sobre prácticas recomendadas de moderación de contenido en épocas de guerra.

Nuestras acciones:

Meta siguió una serie de pasos para abordar los desafíos, incluidos los siguientes:

- **Conformamos un centro de operaciones especiales** con expertos de la empresa disponibles las 24 horas, que cuenta con competencias en idioma ruso y ucraniano, para supervisar la situación y responder al conflicto, que evoluciona con rapidez, en tiempo real.
- **Intensificamos nuestros esfuerzos para abordar contenido infractor**, como violencia e incitación, lenguaje que incita al odio, comportamiento no auténtico coordinado, información errónea y desinformación. Esto incluyó ampliar nuestra capacidad de verificación de datos independiente en ucraniano y ruso en toda la región, y proporcionar ayuda económica adicional a los socios de verificación de datos de Ucrania. También incorporamos cinco nuevos [socios de confianza](#), a quienes convocamos a sesiones de escucha para entender mejor los riesgos relacionados con contenido y desarrollar mecanismos de reporte apropiados.

- **Limitamos la difusión en todo el mundo de contenido de páginas de Facebook y cuentas de Instagram de medios de comunicación controlados por el Estado ruso** para que fuesen más difíciles de encontrar en nuestras apps. Asimismo, comenzamos a etiquetar y limitar publicaciones e historias de Facebook e Instagram con enlaces a sitios web de medios de comunicación bajo el control del Gobierno de Rusia. En la UE y Ucrania, bloqueamos a pedido de los respectivos gobiernos todos los medios de comunicación del Estado ruso.
- **Implementamos funciones de seguridad** centradas en los defensores de derechos humanos y otros grupos vulnerables, por ejemplo, les permitimos bloquear su perfil de Facebook para que nadie acceda a él, eliminamos la opción para que otros puedan ver y explorar listas de amigos y agregamos herramientas en Messenger para proteger a las personas de ataques.
- **Garantizamos que las cuentas de los defensores de derechos humanos y de medios de comunicación independientes** recibieran mayor protección contra acoso o robo de identidad, y que sus voces no se restringieran indebidamente mediante nuestro programa de verificación. Seguimos trabajando en ello.
- **Ayudamos a predecir el flujo de refugiados** gracias a que nuestro equipo de Data for Good comparte conjuntos de datos que preservan la privacidad con socios de confianza.
- **Ayudamos a la población ucraniana** y a otras personas en la región a consultar información confiable de organismos de la ONU locales y sociedades de la Cruz Roja poniendo temporalmente a disposición la ayuda comunitaria como recurso central en Facebook.

Irán

En septiembre de 2022, Mahsa Amini, de 22 años, fue arrestada por la patrulla de orientación iraní por llevar mal puesto su hiyab. Murió bajo custodia policial, lo que desató manifestaciones generalizadas en Irán y en todo el mundo, que aún continúan.

En respuesta a las manifestaciones, las autoridades iraníes impusieron restricciones agresivas a la libertad de expresión y la libertad de reunión, y limitaron el uso de internet y de apps como Instagram y WhatsApp, que la población iraní utiliza ampliamente.

Meta actuó con rapidez y dio una serie de pasos para lograr que las personas se mantuvieran conectadas y seguras en nuestras apps. Esto incluyó reunir a un equipo de personas que hablan farsi, cuya labor central era abordar los problemas derivados de las manifestaciones. El equipo se aseguró de que estuviéramos aplicando nuestras políticas correctamente y protegiendo a nuestros usuarios. Seguimos supervisando la situación de cerca.

Nuestras acciones: ayudar a los manifestantes a dar visibilidad a la situación

La población iraní usa mucho Instagram para dar a conocer las manifestaciones y la realidad de lo que sucede en el lugar. En Instagram, se utilizaron hashtags relacionados con las manifestaciones en Irán más de 160 millones de veces durante los primeros cinco meses de dichas protestas. Los manifestantes compartían sus videos e imágenes de Instagram con medios de comunicación internacionales, ya que muchos de ellos no podían denunciar desde Irán.

- **Ayudar a las personas a conectarse:** En 2022, lanzamos Instagram Lite en Irán para que las personas pudieran acceder a Instagram cuando tienen poca banda ancha. Asimismo, WhatsApp lanzó un servidor proxy que permite a las personas conectarse cuando se bloquea o se corta la conexión a internet, lo que era de particular importancia para las comunidades de Irán.
- **Proteger a activistas y defensores de los derechos humanos:** Las cuentas de medios sociales de activistas y periodistas suelen ser blanco de ataques de bullying, acoso, robo de identidad e intentos de hackeo durante manifestaciones. Trabajamos con activistas iraníes para proteger sus cuentas y ayudarlos a dar a conocer abusos contra los derechos humanos.

10 Nuestro enfoque para los derechos humanos y el desarrollo responsable de productos



Metaverso

En 2021, Meta creó [XR Programs and Research Fund](#), una inversión de dos años por un valor de 50 millones de USD en programas e investigaciones externas con socios del sector, grupos de derechos civiles y humanos, gobiernos, organizaciones sin fines de lucro e instituciones académicas. Estas inversiones se centran en áreas que Meta debe conocer bien para materializar los posibles beneficios del metaverso. Todas las propuestas deben concordar con el trabajo de Meta en una o varias de las cuatro áreas de prioridad normativa que se detallan debajo:

- Oportunidad económica: cómo podemos ofrecer a las personas más opciones y mantener una economía digital próspera.
- Privacidad: cómo podemos crear nuestros productos incorporando la transparencia y controles significativos.
- Seguridad e integridad: cómo podemos proteger a las personas en nuestras plataformas y darles herramientas para que tomen medidas u obtengan ayuda si ven o experimentan algo que las incomoda.
- Igualdad e inclusión: cómo podemos asegurarnos de que el diseño de estas tecnologías sea inclusivo y accesible para todos.

Los socios que forman parte de esta iniciativa incluyen universidades e instituciones de investigación, organizaciones internacionales, grupos de derechos civiles y ONG procedentes de comunidades subrepresentadas de cada región del mundo o que trabajen con ellas.

Inteligencia artificial

Creemos que la inteligencia artificial (IA), desarrollada e implementada con responsabilidad, puede resultar una herramienta eficaz para impulsar los derechos humanos. La IA desempeña un importante papel en cuanto a que nos permite abordar muchos de los impactos adversos para los derechos humanos que mencionamos en este informe. Por ejemplo, el poder de la IA permitió ampliar la detección y revisión de contenido problemático en una creciente variedad de idiomas, mejoró la identificación de contactos e interacciones potencialmente no deseados, mejoró la accesibilidad por medio de la generación de subtítulos automáticos e impulsó varias de nuestras [iniciativas de Data for Good](#).

Al mismo tiempo, el desarrollo y uso de la IA supone nuevos riesgos para los derechos humanos. Entre otros posibles riesgos, los modelos de IA pueden reflejar sesgos problemáticos o tener impactos discriminatorios, así como generar contenido problemático o ejercer una aplicación desmedida de políticas de un modo que afecta de manera negativa la libertad de expresión.

Estamos [comprometidos a abordar estos asuntos](#) de forma abierta y colaborativa. Por ejemplo, los datos demográficos son esenciales para evaluar si un servicio, una app o un proceso [trata a todos los grupos con igualdad](#). Sin embargo, recopilar estos datos conlleva un proceso delicado y genera importantes inquietudes respecto de cómo proteger la privacidad de las personas.

El eje de nuestro enfoque son la responsabilidad y los derechos humanos. La IA se incorpora en nuestra [Política corporativa de derechos humanos](#). Asimismo, reconocemos la importancia de los [Principios de la OCDE sobre Inteligencia Artificial](#), que hacen referencia explícita a los derechos humanos y que el G20 adoptó y apoyó abiertamente. Esperamos fomentar un diálogo más colaborativo y transparente entre las diversas disciplinas y públicos sobre el camino a seguir respecto de estos asuntos tan importantes.

Creamos un equipo de [IA responsable](#) especializado e interdisciplinario cuya tarea es garantizar que la IA en Meta redunde en beneficio de las personas y la sociedad. El equipo incluye especialistas en ética, científicos sociales y políticos, expertos en derechos y políticas, investigadores e ingenieros. El equipo de IA responsable trabaja en estrecha colaboración con el equipo de derechos humanos a fin de asegurar que nuestros sistemas de aprendizaje automático se diseñen y usen de modo responsable.

También fuimos pioneros en un enfoque de [tarjetas de sistemas de IA](#) para proporcionar transparencia estandarizada sobre cómo funcionan estos sistemas, no solo los modelos individuales. Para comenzar, realizamos una prueba piloto centrada en la [clasificación en el feed de Instagram](#).



Cinco pilares de la IA responsable

La IA responsable en Meta se basa en cinco pilares:

- Privacidad y seguridad
- Igualdad e inclusión
- Solidez y protección
- Transparencia y control
- Responsabilidad y gobernanza

Avanzamos en cuanto a abordar muchos de estos asuntos, pero aún queda mucho por hacer. En 2022, trabajamos para:

- **Ampliar los idiomas admitidos**
 - Nuestra iniciativa de código abierto [No Language Left Behind](#) se basa en la IA para ofrecer traducciones de alta calidad de manera directa en 200 idiomas.
 - Nuestro modelo de reconocimiento de [discurso multilingüe masivo](#) permite transcribir el discurso oral a texto y pasar el texto a discurso oral en 1.107 idiomas, y admite la identificación de más de 4.000 idiomas. A algunos de los idiomas admitidos, como el tayuyo, solo los usa un pequeño porcentaje de personas, y para muchos de ellos no existe tecnología de discurso oral previa. Compartimos públicamente los modelos y el código.
 - [Wikipedia se asoció con Meta AI](#) para verificar citas con el programa de IA de código abierto de Meta, [Sphere](#).
- **Abordar los sesgos**
 - En nuestra determinación por garantizar que los sistemas de IA traten a todos con igualdad, introdujimos y [publicamos con código abierto varios conjuntos de datos](#) y modelos responsables para acortar esta brecha de maneras que protejan la privacidad.
 - Entrenamos un [modelo de IA para reducir los sesgos demográficos](#) en el texto (perturbador de texto demográfico), lo que puede contribuir a terminar con las asociaciones estereotípicas presentes en los datos de procesamiento de lenguajes naturales.
 - Introdujimos y publicamos con código abierto dos nuevos conjuntos de datos para medir la igualdad y mitigar posibles sesgos en modelos de NLP. Estos conjuntos de datos incluyen una representación más integral de diferentes dimensiones demográficas para medir la igualdad en estos modelos, incluidos términos que hacen referencia a identidad de género, edad, raza y discapacidades.
- **Entender mejor las asociaciones de contenido problemáticas**
 - Reunimos a un equipo interdisciplinario, que incluye miembros de nuestros equipos de derechos civiles, derechos humanos, ingeniería, IA responsable, investigación de IA, políticas y productos, para entender mejor las asociaciones de contenido problemáticas en varios de nuestros sistemas de extremo a extremo. Trabajamos para implementar medidas de mitigación técnicas para reducir la posibilidad de que haya asociaciones de contenido problemáticas en nuestras apps que usan modelos de IA.

- **Evaluar la igualdad**

- Una manera en la que abordamos la igualdad en el ámbito de la IA es por medio de la creación y distribución de conjuntos de datos más diversos. Asimismo, mediante nuestro proyecto [Casual Conversations v2](#), diseñamos también una base de datos amplia basada en consentimiento para medir la solidez y el sesgo algorítmico.
- Meta AI Research comenzó a explorar [nuevos indicadores de igualdad públicos](#) para evaluar de manera cuantitativa tres tipos de daños y sesgos potenciales bien documentados en modelos de visión artificial. Estos indicadores de igualdad complementan los enfoques actuales respecto del desarrollo de IA responsable, como la documentación de datos y modelos. Asimismo, se diseñaron específicamente para adaptarse y evolucionar a medida que progresa la investigación y emergen nuevos enfoques.

También logramos [avances en SEER](#) (SElf-SupERvised), el modelo de visión artificial autosupervisado de Meta AI Research cuyo objetivo es mejorar los resultados de conjuntos de imágenes diversas, sin necesidad de realizar la cuidadosa selección y el riguroso etiquetado de datos que conlleva el entrenamiento de visión artificial tradicional.





Las normas y los principios de derechos humanos siguen siendo la mejor herramienta de la humanidad para expresar nuestros valores universales, así como la base más sólida sobre la cual crear instituciones de gobernanza legítimas. Los modelos de gobernanza de internet lograron un enorme progreso durante los últimos años. Como este informe demuestra, trabajamos durante todo 2022 para seguir logrando avances con nuestros modelos de gobernanza y hacer el bien en nuestra [Política corporativa de derechos humanos](#) y nuestro compromiso con los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#).

Esta labor continúa, así como también nuestro compromiso con el progreso. Esperamos implementar muchos de los mecanismos de responsabilidad avanzados que se analizan en este informe. Por ejemplo, seguimos trabajando para permitir que las personas participen en medidas de gobernanza por medio de asambleas y foros comunitarios.

Asimismo, seguimos trabajando para que los derechos humanos sean el eje de nuestro conocimiento e implementación de la ciencia de inteligencia artificial, que avanza exponencialmente. Si bien tal vez nos esperen muchos desafíos en el camino y haya mucho por hacer, nos enorgullecen nuestras iniciativas de amplio alcance con las que resolvemos los problemas más complejos de los medios sociales y las tecnologías de IA.

Seguiremos basándonos en la debida diligencia en materia de derechos humanos para precisar mejor cuáles son los riesgos para los derechos humanos más importantes que enfrentan las personas que usan nuestros servicios y para elaborar una política que aborde dichos riesgos.

Operamos en un entorno regulatorio complejo y fracturado. La naturaleza sin fronteras y gratuita del internet que muchos de nosotros usamos a diario se encuentra amenazada. Es fundamental que trabajemos con gobiernos, las Naciones Unidas y la sociedad civil para garantizar que los marcos normativos se desarrollen sobre la base de valores y principios de derechos humanos. Con plena conciencia de nuestras responsabilidades con los derechos humanos, nos esforzamos por fomentar estos principios en nuestro trabajo cotidiano en beneficio de nuestros usuarios y de la sociedad en general.

