



∞ Meta

Informe sobre derechos humanos

Información valiosa y medidas
2023



Tabla de contenidos

Acerca de este informe	3
Resumen ejecutivo	5
Prólogo	8
La IA en el contexto de los derechos humanos	9
Adoptar un enfoque abierto	9
Abordar la problemática de las respuestas potencialmente dañinas de la IA generativa	11
Gestión de riesgos para los derechos humanos	13
Diligencia debida en el día a día	13
Política de contenido	16
Desarrollo de productos	17
Cómo nos preparamos para las situaciones críticas y las crisis mundiales y cómo respondemos a ellas	19
Cómo se prepara Meta para las elecciones	19
Brasil	20
Pakistán	22
Respuesta a las crisis mundiales	24
Sudán	24
Nagorno Karabaj	26
Migración en América Central y América del Sur	28
Evaluaciones sobre el impacto en los derechos humanos	29
Israel y Palestina: actualización sobre nuestras medidas	29
Medidas temporales de políticas	30
Medidas temporales de productos	31
Cifrado de extremo a extremo: Actualización sobre nuestras medidas	32
Temas destacados	33
Combatir las amenazas de seguridad	33
Seguridad infantil y juvenil	34
Explotación infantil	34
Seguridad y bienestar juvenil	36
Participación de partes interesadas	37
Grupos marginados y defensores de los derechos humanos	37
Socios de confianza	39
Caso práctico: mostrar los nuevos daños en Haití	40
Caso práctico: mejorar los reportes en Etiopía	40
Caso práctico: mitigar los conflictos interreligiosos en Bangladesh	40
Organizaciones internacionales y de varias partes interesadas	41
Transparencia y reparación	44
Responder a solicitudes gubernamentales	45
Consejo asesor de contenido	50
Consideraciones para el futuro	53
Anexo	55
Cómo Meta gestiona y rige el trabajo relativo a los derechos humanos	55
Capacitación en materia de derechos humanos para empleados de Meta	55
Enlaces a los informes mencionados	56



Acerca de este informe

Este es nuestro tercer informe anual sobre derechos humanos. Se basa en el trabajo realizado por Meta y refleja el progreso conseguido respecto 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](#). Comprende las medidas que implementamos desde el 1 de enero hasta el 31 de diciembre de 2023. Incluimos información sobre los servicios y productos de Meta, como Facebook, Messenger, Instagram, WhatsApp, Threads y Reality Labs.

Los derechos humanos son un tema fundamental tanto para la empresa como para nuestras partes interesadas. El contenido de este informe refleja los comentarios que recibimos de nuestras partes interesadas luego de la publicación de nuestro [Informe sobre derechos humanos de 2022](#) y se basa en nuestra [evaluación integral de riesgos significativos para los derechos humanos](#), que se encuentra resumida en el informe de 2022. El propósito de la evaluación era identificar y priorizar los impactos¹ potencialmente adversos para los derechos humanos más importantes, que afectan a usuarios y a otras personas sobre quienes repercuten las medidas que adoptamos en la empresa en todo el mundo. Este informe destaca la manera en que Meta gestionó en sus operaciones internacionales los siguientes ocho derechos humanos destacados:

¹ El término "impacto adverso para los derechos humanos" concuerda con lo establecido en los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas y se refiere al impacto que sufre una persona cuando una medida reduce o anula su capacidad de gozar de sus derechos humanos.

Principales ocho potenciales derechos humanos destacados según nuestra evaluación integral de riesgos significativos para los derechos humanos

Derecho humano destacado

Sección del informe

Libertad de opinión y expresión

Política de contenido, Cómo se prepara Meta para las elecciones, Respuesta a las crisis mundiales, Israel y Palestina: actualización sobre nuestras medidas, Seguridad infantil y juvenil, Responder a solicitudes gubernamentales, Consejo asesor de contenido

Privacidad

Política de contenido, Cifrado de extremo a extremo: actualización sobre nuestras medidas, Combatir las amenazas de seguridad, Seguridad infantil y juvenil, Responder a solicitudes gubernamentales

Igualdad y no discriminación

La IA en el contexto de los derechos humanos, Desarrollo de productos, Israel y Palestina: actualización sobre nuestras medidas

Vida, libertad y seguridad de la persona

Cómo se prepara Meta para las elecciones, Respuesta a las crisis mundiales

Interés superior del niño

Política de contenido, Seguridad infantil y juvenil, Participación de partes interesadas

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

Cómo se prepara Meta para las elecciones: Brasil, Pakistán

Libertad de reunión y asociación

Cifrado de extremo a extremo: actualización sobre nuestras medidas, Respuesta a las crisis mundiales: Sudán

Derecho a la salud

Respuesta a las crisis mundiales: Sudán



En la parte superior de las secciones de este informe, se encuentran los riesgos significativos para los derechos humanos.

Trabajamos con diferentes partes externas e internas y nos esforzamos por proporcionar un panorama representativo de nuestro trabajo en diferentes equipos de todo el mundo. El informe tiene por objeto proporcionar información general sobre los problemas en materia de derechos humanos que enfrentamos como empresa y, a la vez, guiar al lector para que pueda explorar contenido más detallado si lo necesita.

Este informe complementa el último [Informe de prácticas comerciales responsables de Meta](#). [Informamos](#) por separado nuestras iniciativas para identificar y mitigar los riesgos en materia de esclavitud moderna y trata de personas en nuestras operaciones comerciales y cadenas de suministro. Se pueden consultar enlaces a otras divulgaciones de Meta en el [Anexo](#) de este informe.

Nuestra Política corporativa de derechos humanos 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

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.





Resumen ejecutivo



Con este informe sobre derechos humanos buscamos brindar información valiosa sobre las políticas, procesos y normas que Meta utiliza para gestionar los riesgos de derechos humanos de miles de millones de personas, a escala y en tiempo real. Nuestro objetivo es transmitir cómo nos esforzamos por respetar los derechos de los usuarios y otras comunidades, en nuestro trabajo y en los sistemas que utilizamos cada día, desde la manera en que respondemos a las crisis y planificamos para las elecciones hasta la forma en que respondemos a las exigencias gubernamentales.

El informe de este año comprende nuestras actividades del año calendario 2023 y refleja los comentarios de las partes interesadas con respecto al [informe sobre derechos humanos del año pasado](#). El informe de este año detalla cómo Meta gestionó en sus operaciones internacionales

los siguientes ocho derechos humanos destacados: libertad de opinión y expresión; privacidad; igualdad y no discriminación; vida, libertad y seguridad de la persona; interés superior del niño; derecho a la participación pública, al voto y a presentarse a una candidatura; libertad de reunión y asociación; y el derecho a la salud.

Los servicios y las apps de Meta seguirán utilizándose para hacer el bien: para ayudar a construir una comunidad, fomentar la innovación y movilizar los recursos para apoyar la ayuda humanitaria, entre otras iniciativas. A medida que surgieron desafíos, nuestra [Política corporativa de derechos humanos](#) orientó nuestras medidas para abordar temas como la utilización de nuestros servicios en países en conflicto, la ciberseguridad y las solicitudes gubernamentales de datos desproporcionadas.



Seguimos esforzándonos por tener presentes a los derechos humanos en la manera en que entendemos la inteligencia artificial (IA) y la desarrollamos. Nuestros productos de IA generativa permiten a las personas promover los derechos humanos de formas novedosas, pero reconocemos que estos sistemas no son perfectos y presentan oportunidades de mejoras continuas. Para mitigar los riesgos potenciales, desarrollamos una [Política de uso aceptable](#), incorporada al [Acuerdo de licencia comunitaria](#) para nuestro modelo abierto Llama 2, y proporcionamos herramientas de seguridad para ayudar a los desarrolladores a crear con responsabilidad. Creemos que un enfoque abierto permite impulsar la innovación a través de la iteración y favorece la creación de productos mejores y más seguros gracias a las opiniones, el escrutinio, el desarrollo y las mitigaciones constantes.

El trabajo proactivo con las partes interesadas es fundamental para nuestro enfoque sobre los derechos humanos y está incorporado en nuestra Política corporativa de derechos humanos. El desarrollo de los servicios, las políticas de contenido, los procesos de moderación y las Normas comunitarias [se basan](#) en información valiosa obtenida de una amplia variedad de organizaciones de la sociedad civil, defensores de los derechos humanos, grupos marginados, organizaciones internacionales, inversores, anunciantes, usuarios y [socios de confianza](#). Estos grupos abarcan todo el espectro político e ideológico a fin de proporcionar diferentes puntos de vista. En enero de 2023, nuestra red de socios de confianza incluía a más de 400 organizaciones no gubernamentales, agencias humanitarias, organizaciones de defensa de los derechos humanos e investigadores de 113 países de todo el mundo.

El año 2023 marcó el décimo aniversario de la incorporación de Meta a la [Global Network Initiative](#) (GNI). Aplicamos los

principios de la GNI (y tenemos la responsabilidad de cumplirlos) en lo relativo a la manera en que las empresas de tecnología deben respetar los derechos de privacidad y libertad de expresión de sus usuarios a la hora de abordar solicitudes y restricciones gubernamentales. En el informe, describimos la manera en que la GNI fue la base de nuestra respuesta a determinadas solicitudes gubernamentales, incluidas restricciones de contenido y acceso a datos de usuarios.

También trabajamos en integrar los principios de derechos humanos para prepararnos para los problemas y situaciones críticas de 2023 y responder a ellos. El informe abarca la manera en que nos preparamos para enfrentar desafíos intensificados en países de alto riesgo, acontecimientos críticos planificados (como elecciones) y nuestra respuesta coordinada a los eventos de emergencia. En 2023, Meta posibilitó la preparación para más de 130 elecciones en todo el mundo. En este informe, incluimos nuestra respuesta a los conflictos en Sudán y Nagorno Karabaj, así como a los problemas de derechos humanos relacionados con la migración en América Central y América del Sur. Nuestras iniciativas de respuesta a crisis involucraron a diferentes expertos en áreas como política de contenido, operaciones internacionales, derechos humanos y producto, así como otros expertos internos que evaluaron y respondieron a estas situaciones, basándose en lo establecido en el [protocolo de la política de crisis](#). También describimos la manera en que pusimos en práctica las recomendaciones de procesos anteriores de diligencia debida en materia de derechos humanos respecto de [Israel y Palestina](#) y las medidas que adoptamos en el marco del último conflicto, así como también una actualización sobre el cifrado de extremo a extremo.



Reforzamos nuestro compromiso con la transparencia y la reparación como principios rectores de la gestión de riesgos para los derechos humanos. Seguimos trabajando con el [Consejo asesor de contenido](#), una iniciativa líder en el sector creada para ayudar a Meta a abordar algunas de las cuestiones más complejas en torno a la libertad de expresión y a la seguridad online. El año 2023 fue un año histórico para el Consejo asesor de contenido, que superó su objetivo de 50 decisiones por año, más del triple de las decisiones tomadas en 2022.

Además, mejoramos nuestra experiencia de usuario en Facebook y en Instagram para el contenido que Meta restringe con motivo de solicitudes legales gubernamentales y por parte de tribunales. En la mayoría de los casos, mediante nuestras notificaciones informamos a los usuarios qué autoridad estatal envió la solicitud que generó la restricción y el país en que se restringió el contenido.

En 2023, enviamos nuestra primera serie de solicitudes de eliminación de Austria, Vietnam, Singapur, India y México a la base de datos de [Lumen](#). Este proyecto de investigación independiente, a cargo del Berkman Klein Center for Internet & Society de la Universidad de Harvard, permite a los investigadores analizar las solicitudes de eliminación de contenido online efectuadas por agentes privados y gubernamentales. Esto representa un avance a la hora de posibilitar que la comunidad internacional analice y defienda los derechos digitales de los usuarios de internet y prepare informes al respecto.





Prólogo

A medida que el mundo evoluciona, continuamos enfrentándonos a desafíos nuevos y complejos. La manera en que respondemos a estos desafíos se basa en nuestra [Política corporativa de derechos humanos](#) y en nuestro compromiso con los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#) y el [Pacto Mundial de las Naciones Unidas](#). Y tenemos un sistema de controles y equilibrios.

El [Consejo asesor de contenido](#) ofrece una supervisión independiente a cargo de expertos para garantizar la toma de decisiones justas y equilibradas, en especial, en lo relativo a cuestiones vinculadas a la libertad de expresión.

Nuestras políticas y procedimientos también evolucionan a medida que la tecnología y la sociedad cambian. Consideramos la opinión de una amplia gama de partes interesadas haciendo énfasis en las múltiples dimensiones de la diversidad, incluida la diversidad de puntos de vista. A la hora de revisar estas políticas y procedimientos, pedimos la opinión de empleados, expertos y defensores de los derechos humanos, así como de grupos marginados, organizaciones internacionales y una amplia variedad de grupos de la sociedad civil.

Utilizamos diferentes protocolos y herramientas de evaluación en toda la empresa para identificar, evitar y mitigar los riesgos relacionados con los derechos humanos que puedan vincularse con los productos, las políticas y las operaciones de Meta en todo el mundo. Brindamos información sobre evaluaciones anteriores con el objetivo de compartir el avance y ofrecer canales abiertos para recibir comentarios.

En 2023, aplicamos nuestras herramientas de gestión de riesgos al responder a situaciones de emergencia, incluidos conflictos y guerras en África y Oriente Medio, así como para planificar

cuestiones relacionadas con acontecimientos mundiales previsible, como la cantidad considerable de procesos electorales. En este sentido, nuestras políticas son dinámicas a fin de garantizar que podamos responder de forma adecuada y personalizada a cada situación al tiempo que logramos un equilibrio entre los intereses contrapuestos y las consideraciones en materia de derechos humanos.

Meta es pionera en el desarrollo de inteligencia artificial (IA) hace más de una década. Sabemos que el progreso y la responsabilidad deben ir de la mano. Las herramientas de IA generativa ofrecen amplias oportunidades, y creemos que es tanto posible como necesario que estas tecnologías se desarrollen de un modo que respete los derechos humanos.

Un mercado digital sólido y abierto es esencial para ayudar a las personas, la sociedad y los países a prosperar y consolidar la comunidad. Hoy más que nunca, necesitamos contar con tecnología que nos ayude a unir más al mundo. Nos esforzamos por desempeñar nuestra función con una perspectiva de derechos humanos firmemente arraigada en nuestras iniciativas.

Firmado por:



Nick Clegg
Presidente de Asuntos Internacionales



Jennifer Newstead
Directora ejecutiva de Asuntos Legales



La IA en el contexto de los derechos humanos

Riesgos significativos relevantes:

- Libertad de opinión y expresión
- Igualdad y no discriminación
- Privacidad

El rápido avance de la inteligencia artificial (IA), en especial, la [IA generativa](#), fue el centro de la atención mundial en 2023. Debido a la mayor atención y a las posibles oportunidades y desafíos que presenta la IA, ofrecemos un análisis detallado del trabajo que hacemos en el campo de la IA desde una perspectiva de derechos humanos.

En 2023, lanzamos nuestro modelo de lenguaje grande abierto [Llama 2](#), el [asistente Meta AI](#), el modelo de generación de imágenes [Emu](#) y las [herramientas creativas en la app](#).

A medida que las herramientas de generación de imágenes y los chatbots con IA sofisticados comenzaron a utilizarse de forma generalizada, también surgió la necesidad de supervisión por parte de organismos públicos y regulatorios. Por ejemplo, la Unión Europea dictó su histórica [Ley de IA](#), mientras que Estados Unidos obtuvo una serie de [compromisos voluntarios de empresas de IA con la Casa Blanca](#), como Meta y otros desarrolladores de IA.

Nos comprometemos a desarrollar e implementar la IA con responsabilidad y, a la vez, mitigar cualquier potencial impacto negativo en materia de derechos humanos. La IA forma parte de nuestra [Política corporativa de derechos humanos](#), que reconoce la importancia de los [principios sobre inteligencia artificial de la OCDE](#) que el G20 adoptó y apoyó ampliamente.

Creemos que la IA responsable puede ser una potente herramienta a la hora de impulsar los derechos humanos.

Creemos que la IA responsable puede ser una potente herramienta a la hora de defender los derechos humanos. Nuestros productos de IA generativa permiten a las personas ejercer su libertad de expresión de formas novedosas, aumentan el acceso a la información y a la educación y mejoran la accesibilidad. Por ejemplo, los subtítulos generados con IA y el reconocimiento de imágenes de los lentes inteligentes Ray-Ban Meta mejoran la accesibilidad de las personas con discapacidades.

Aprovechamos la IA para detectar contenido potencialmente infractor e interacciones no deseadas y responder a ellos rápidamente en una cantidad cada vez mayor de idiomas a fin de evitar que estas situaciones se agraven. Muchas de nuestras [iniciativas de Data for Good](#) usan la IA para apoyar las respuestas humanitarias a situaciones de emergencia y facilitar el trabajo en materia de salud pública.

Adoptar un enfoque abierto

Meta procura adoptar un enfoque responsable en el desarrollo y la implementación de modelos y productos de IA generativa.

Creemos que un enfoque abierto permite impulsar la innovación a través de la iteración y favorece la creación de productos mejores y más seguros gracias a las opiniones, el escrutinio, el desarrollo y las mitigaciones constantes por parte de la comunidad. Además permite lo siguiente:

- Promover la igualdad y aceptar diferentes puntos de vista al permitir que los titulares de derechos afectados identifiquen mejor los sesgos potenciales.
- Disminuir los obstáculos para la innovación, lo que genera beneficios económicos.
- Posibilitar que las comunidades internacionales personalicen con mayor facilidad los avances en IA para contextos e idiomas específicos.

Sin embargo, reconocemos que la adopción de un enfoque abierto también puede presentar riesgos potenciales para los derechos humanos, lo que incluye la posibilidad de que los desarrolladores desestimen las instrucciones de uso responsable de Meta e implementen modelos de formas que no sean seguras. Para tratar de mitigar estos riesgos, realizamos pruebas de seguridad y desarrollamos una [Política de uso aceptable](#), incorporada al [Acuerdo de licencia comunitaria](#) para nuestros modelos Llama. También compartimos una [Guía de uso responsable](#) detallada y proporcionamos herramientas de seguridad a través de nuestro proyecto [Purple Llama](#), incluido [Llama Guard](#), para ayudar a los desarrolladores a crear de modo responsable con estos modelos fundacionales disponibles para el público en general e implementarlos de forma segura en sus propios casos de uso.

Creemos que un enfoque abierto permite impulsar la innovación a través de la iteración y favorece la creación de productos mejores y más seguros gracias a las opiniones, el escrutinio, el desarrollo y las mitigaciones constantes por parte de la comunidad.

Implementar la IA de forma segura es una responsabilidad compartida en todo el ecosistema, y es por eso que hace muchos años colaboramos con las organizaciones que están trabajando para crear una IA confiable y segura. Por ejemplo:

- Venimos trabajando con [MLCommons](#) y una serie de socios internacionales para crear [puntos de referencia de responsabilidad](#) de maneras que beneficien a la comunidad de código abierto.
- Participamos en iniciativas multisectoriales como la [Partnership on AI](#) para abordar cuestiones que van desde la implementación responsable de los modelos fundacionales hasta las maneras más adecuadas de identificar el contenido sintético.
- Cofundamos la [AI Alliance](#), una coalición de empresas, académicos, colaboradores y Gobiernos que trabajan para desarrollar herramientas que permitan crear un ecosistema de IA abierto y seguro.

Consultamos a partes interesadas en materia de derechos humanos para ayudarlas a entender el uso que hacemos de la IA y para que nos den su opinión al respecto. En el marco de estas iniciativas, colaboramos con el [trabajo en IA generativa](#) que lleva a cabo el [proyecto B-Tech](#) de la Oficina de Derechos Humanos de las Naciones Unidas, y brindamos información para la elaboración de [tres documentos fundacionales](#) basados en los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas. También trabajamos con un asesor de la Oficina de la Relatora Especial de la ONU sobre la libertad de opinión y expresión y la UNESCO para dar a conocer los conceptos fundamentales y los riesgos de la IA, y posibilitamos debates fundados para entender las preocupaciones. Además, [realizamos un foro de la comunidad consultivo](#), en asociación con la Universidad de Stanford, que incorporó preguntas sobre derechos humanos para ayudar a Meta y a otras empresas, investigadores y Gobiernos a tomar decisiones sobre los chatbots con IA generativa. Entre las [conclusiones](#) se destacaron la importancia de contar con un marco de derechos humanos y de incorporar información proveniente de fuentes relacionadas con los derechos humanos.

Abordar la problemática de las respuestas potencialmente dañinas de la IA generativa

Somos conscientes de que la tecnología de IA generativa podría generar respuestas que pueden, por ejemplo, producir contenido potencialmente dañino, ofensivo o discriminatorio, así como reforzar sesgos, presentar información inexacta o plantear consideraciones en materia de privacidad. También sabemos que personas malintencionadas podrían hacer un uso indebido de los modelos y las herramientas que lanzamos de maneras deliberadamente dañinas. Implementamos una serie de pasos para mitigar estos riesgos en nuestros modelos fundacionales, incluidos Llama y Emu. Creemos que adoptar un enfoque abierto

e iterativo nos ayudará a mejorar en tiempo real en el caso de que se planteen estos problemas.

El proyecto UN B-Tech [destaca](#) una serie de áreas de riesgo relevantes en relación con varios derechos reconocidos a nivel internacional. Tomamos esta guía y los [riesgos significativos para los derechos humanos](#) de Meta como base de nuestro trabajo.

En el caso de nuestros modelos fundacionales [Llama 2](#) y [Emu](#), estas mitigaciones buscan en parte abordar los potenciales riesgos significativos para los derechos humanos al abordar los riesgos en el entrenamiento, analizar las asociaciones sensibles en función del contexto y realizar evaluaciones y ajustes de seguridad.

A medida que seguimos desarrollando nuestros modelos fundacionales, compartimos [información detallada](#) sobre nuestro enfoque responsable de forma pública.

Además de nuestras iniciativas proactivas para mitigar los riesgos asociados con las respuestas generadas con IA potencialmente dañinas, también trabajamos para abordar estas respuestas a medida que tomamos conocimiento de ellas.



Imagen generada con Meta AI



En los días siguientes a los atentados terroristas del 7 de octubre en Israel, diferentes partes interesadas nos notificaron de varios problemas en nuestras herramientas, los cuales reconocimos y procuramos corregir rápidamente. Los principales problemas se relacionaban con traducciones automáticas en las que se agregó la palabra "terrorista" a los datos personales traducidos de los perfiles de Instagram de algunos usuarios y con stickers generados con IA que mostraban a los palestinos como violentos.

En cuanto tomamos conocimiento de estos problemas, nuestros equipos de ingeniería iniciaron una investigación para determinar las causas principales e implementar correcciones. Nuestros equipos identificaron que estos problemas parecían estar relacionados con alucinaciones del modelo y con los datos de entrenamiento utilizados. Ambas problemáticas constituyen desafíos ampliamente documentados relacionados con los productos desarrollados con IA (consulta [este enlace](#) y [este enlace](#)).

Proporcionamos una solución de emergencia para nuestro producto de traducción automática en un plazo de 90 minutos desde que nos enteramos de la existencia del inconveniente. También trabajamos con la mayor velocidad posible para mitigar las asociaciones potencialmente problemáticas en los resultados de imágenes y stickers generados en relación con el 7 de octubre. Por ejemplo, agregamos apellidos conocidos al modelo para evitar las alucinaciones. Además, ajustamos nuestros modelos fundacionales de generación de imágenes para abordar mejor una amplia gama de asociaciones potencialmente problemáticas en relación con los conflictos e implementamos dichos ajustes en todos nuestros productos.

Consulta la sección [Israel y Palestina: actualización sobre nuestras medidas](#) para obtener más información.

Esperamos seguir colaborando con una amplia variedad de titulares de derechos con diferentes puntos de vista a medida que nos esforzamos por poner a la seguridad, la responsabilidad y los derechos humanos en el centro de nuestro trabajo a la hora de desarrollar e implementar productos de IA generativa. También adoptaremos las medidas adecuadas para cumplir con las obligaciones legales en esta área.

Gestión de riesgos para los derechos humanos

Todos los días, nos esforzamos activamente por convertir los principios de derechos humanos en acciones significativas. La [Política corporativa de derechos humanos de Meta](#) es la base de nuestra labor. Nuestro compromiso y enfoque se basan en los [Principios Rectores sobre las Empresas](#) y los [Derechos Humanos de las Naciones Unidas](#), junto con las normas internacionales y regionales en materia de derechos humanos detalladas en nuestra Política corporativa de derechos humanos.

Nuestro trabajo incluye, entre otras iniciativas, implementar procesos de diligencia debida, fomentar la [participación de las partes interesadas](#), ofrecer [asesoramiento en materia de productos](#), [proporcionar capacitación sobre derechos humanos](#) y brindar asesoramiento basado en datos en lo relativo a todos los aspectos del negocio. En este informe, nos centramos en los diferentes tipos de procesos de diligencia debida en materia de derechos humanos que implementamos en Meta. Para implementar los procesos de diligencia debida, aplicamos una serie de protocolos y herramientas de evaluación para identificar, evitar y mitigar los riesgos que podrían vincularse con nuestros productos, políticas y operaciones en todo el mundo.



Este trabajo refleja la manera en que entendemos la regulación, junto con nuestros compromisos con la [Global Network Initiative \(GNI\)](#).

El marco regulatorio está cambiando y presenta implicancias en materia de derechos humanos, como muestran las últimas tendencias mundiales. Por ejemplo, en 2023, fuimos testigos de la sanción de la Ley de Seguridad Online del Reino Unido y de la entrada en vigencia de la Ley de Servicios Digitales de la Unión Europea, así como de otras leyes de seguridad online en países como Singapur, Taiwán y Uruguay.

La tabla de la página 15 recoge algunas de las medidas que implementamos

en materia de gestión de riesgos para los derechos humanos. También hay más detalles disponibles en nuestros informes de derechos humanos de [2021](#) y [2022](#).

Diligencia debida en el día a día

A través de nuestra labor en materia de diligencia debida en el día a día, muchos equipos diferentes colaboran para integrar los derechos humanos en nuestras actividades continuamente. Por ejemplo, los expertos en derechos humanos y derechos civiles de Meta proporcionan asesoramiento constante para evaluar y mitigar los

riesgos para los derechos humanos en la política de contenido, el desarrollo de productos, la respuesta ante situaciones de emergencia o conflicto y la preparación para las elecciones. Este trabajo a menudo implica abordar cuestiones relacionadas con el lenguaje que incita al odio, información errónea, contenido gráfico que puede documentar abusos de los derechos humanos y medidas que ayuden a proteger a los defensores de los derechos humanos, así como otras cuestiones relacionadas con los derechos humanos.

Con más de 3.190.000 millones de personas en promedio que utilizan al menos una de nuestras apps a diario según datos hasta diciembre de 2023, los impactos potenciales de nuestros productos en los derechos humanos varían considerablemente en función del momento, la ubicación, el contenido y las comunidades afectadas. Por este motivo, priorizamos nuestro trabajo basándonos en la [evaluación integral de riesgos significativos para los derechos humanos](#) de 2022, en la que se analizaron los impactos en todos los derechos humanos reconocidos a nivel internacional. El análisis priorizó los ocho riesgos más significativos utilizando los criterios establecidos en los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#), que se detallan en el gráfico que figura debajo.

A menudo tenemos dificultades respecto de la manera en que priorizamos las medidas que adoptamos en relación con los riesgos potenciales para los derechos humanos. Esta tensión es inherente al marco de los derechos humanos y se pone de manifiesto cuando buscamos conciliar diferentes derechos humanos, en ocasiones contrapuestos. Buscamos priorizar los riesgos más significativos para los derechos humanos, es decir, aquellos con el potencial de provocar los impactos negativos más graves a las partes interesadas.

Buscamos priorizar los riesgos más significativos para los derechos humanos, es decir, aquellos con el potencial de provocar los impactos negativos más graves a las partes interesadas.

La forma en que establecemos prioridades se basa en los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas y nuestra Política corporativa de derechos humanos, a lo que se suma la participación de las partes interesadas en torno a múltiples dimensiones de diversidad. Por ejemplo, es necesario establecer prioridades para permitir el uso de nuestros productos y servicios por parte de categorías vulnerables de personas, como los jóvenes, dentro de la población general amplia y diversa de usuarios. Meta utiliza la [Convención sobre los Derechos del Niño de la ONU](#), en especial, el principio del "Interés superior del niño" como base para las cuestiones relativas a la prestación de servicios a los jóvenes. En ocasiones, esto da lugar a decisiones complejas entre qué impactos potenciales sobre los derechos tienen prioridad. Por ejemplo, reconocer el derecho a la seguridad y el bienestar de los niños al tiempo que se respeta su derecho a buscar información de forma privada sobre temas como política, salud o identidad de género. El [marco del Interés superior del niño](#) fue la base de la adopción de funciones como el [Centro para familias](#) y la [supervisión parental](#).

Nuestro enfoque de la gestión de riesgos para los derechos humanos

Compromisos con los derechos humanos

Implementamos nuestros compromisos con los derechos humanos mediante enfoques establecidos en los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas, incluidos los siguientes:

Aplicación de políticas de derechos humanos

Aplicación de procesos de diligencia debida y divulgación en materia de derechos humanos

Acceso a recursos

Mantenimiento de la supervisión, gobernanza y responsabilidad

Protección de los defensores de los derechos humanos

Priorización de derechos

La evaluación integral de riesgos significativos para los derechos humanos identificó ocho derechos humanos a los que se les debe dar prioridad:

Libertad de opinión y expresión

Privacidad

Igualdad y no discriminación

Vida, libertad y seguridad de la persona

Interés superior del niño

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

Libertad de reunión y asociación

Derecho a la salud

Principales herramientas y ejemplos

Implementamos las siguientes herramientas y procesos para mitigar los riesgos, y se incluyen en este informe algunos ejemplos ilustrativos:

Desarrollo de la política de contenido → Actualización de personas y organizaciones peligrosas

Asesoramiento en materia de producto → IA generativa; Threads

Integración de los derechos humanos en la preparación para las elecciones → Elecciones en Pakistán

Integración de los derechos humanos en las iniciativas de respuesta coordinadas → Priorización de países que necesitan una respuesta activa

Respuesta a crisis → Conflictos en Nagorno-Karabaj y Sudán

Marco de la GNI para solicitudes gubernamentales → Restricciones de contenido en virtud de la legislación local y el acceso a datos del usuario

Consejo asesor de contenido → Entradas sobre casos de referencia, por ejemplo, rehenes en el derecho internacional

Participación de partes interesadas → Sesiones de diseño conjunto para padres y adolescentes

Protección de los defensores de los derechos humanos → Fondo para defensores de derechos humanos



Riesgos significativos relevantes:

- Libertad de opinión y expresión
- Privacidad
- Igualdad y no discriminación
- Vida, libertad y seguridad de la persona

Política de contenido

Como parte de nuestro trabajo básico de diligencia debida, nuestros expertos en derechos humanos respaldan el proceso de desarrollo de políticas de contenido. Esto significa que revisan propuestas para actualizar las políticas de contenido desde la perspectiva de la legislación sobre derechos humanos, teniendo en cuenta la libertad de expresión y la protección contra la discriminación, así como otros derechos humanos. En 2023, esto incluyó la mejora de la [política de personas y organizaciones peligrosas](#). Este trabajo se llevó a cabo en diferentes etapas a lo largo del 2023 y publicamos una [actualización](#) sobre dicho trabajo en enero de 2024.

De conformidad con la política de personas y organizaciones peligrosas, no permitimos la presencia en los servicios de Meta de organizaciones o personas que reivindiquen una misión violenta o que cometan actos de violencia. Nuestro enfoque global se basa en [definiciones exhaustivas](#) de tipos y niveles de organizaciones peligrosas, de acuerdo con lo publicado en nuestro Centro de transparencia. Contamos con un proceso independiente para identificar a personas y organizaciones peligrosas, y tenemos expertos competentes que revisan los casos en base a evidencia corroborada. Nuestros umbrales y definiciones son neutrales respecto a regiones o ideologías y se desarrollaron en consulta con expertos y académicos de todo el mundo. Además, aplicamos medidas contra las organizaciones terroristas extranjeras designadas por EE. UU., traficantes de narcóticos especialmente designados y terroristas mundiales especialmente designados.

De acuerdo con el [Consejo asesor de contenido](#) y las recomendaciones de [Business for Social Responsibility](#) sobre la diligencia debida en materia de los derechos humanos en Israel y Palestina, [actualizamos](#) nuestra política de personas y organizaciones peligrosas de las siguientes formas:

- Actualizamos la política para permitir más discurso social y político, incluido el relacionado con elecciones, resolución de conflictos y ayuda humanitaria y para desastres, sin dejar de cumplir con nuestras obligaciones legales.
- Recibimos comentarios que indicaban que la definición anterior de "elogio de organizaciones y personas peligrosas" era demasiado amplia, por lo que actualizamos la política para hacerla más matizada y proporcionada. Ahora prohibimos la "glorificación" de la violencia o el odio de organizaciones y personas peligrosas, que se define de forma más restrictiva que los "elogios".
- Simplificamos la forma en que evaluamos y clasificamos a los agentes peligrosos según sus vínculos con la violencia y el daño en la vida real.
- Actualizamos el proceso de eliminación para proporcionar criterios más exhaustivos y detallados que deben cumplirse para que se considere la eliminación de una organización o persona peligrosa de la lista. Esta actualización nos permite asegurar que las designaciones reflejen las circunstancias cambiantes.



En septiembre de 2023, también hicimos cambios en nuestra [política de violencia e incitación](#) para perfeccionar el lenguaje sobre violencia de gravedad alta. En marzo de 2023, realizamos cambios en la [política de bullying y acoso](#) para incluir la definición de "figuras públicas" siguiendo la recomendación del Consejo asesor de contenido. En junio de 2023, llevamos a cabo un Foro de políticas sobre contenido violento y gráfico en contextos de abuso de los derechos humanos como respuesta a la recomendación del Consejo asesor de contenido en el [caso del video gráfico de Sudán](#). Consideramos si nuestras políticas sobre contenido violento y gráfico lograban un equilibrio apropiado entre el respeto a la privacidad y dignidad de las víctimas, la libertad de expresión y el bienestar de las comunidades. El análisis indicó que el equilibrio era apropiado, por lo que no cambiamos las políticas.

Con el fin de contribuir a la transparencia, los cambios en las políticas se reflejan en las [Normas comunitarias](#) y se pueden consultar en el registro de cambios correspondiente y en las [actas públicas del Foro de políticas](#).

Además, actualizamos el [sistema de penalizaciones](#) en 2023. El sistema nuevo redujo la cantidad de períodos de restricción y tiene el objetivo de asegurar la proporcionalidad de las respuestas a las infracciones de políticas. Seguiremos aplicando restricciones a las cuentas de infractores reincidentes, normalmente desde la séptima infracción, tras el envío de suficientes advertencias y explicaciones para ayudar a la persona a entender por qué se eliminó su contenido. En el caso de infracciones más graves, como la publicación de contenido que contenga terrorismo, explotación infantil, trata de personas,

promoción del suicidio, explotación sexual, venta de drogas sin fines medicinales o promoción de personas y organizaciones peligrosas, seguiremos aplicando consecuencias inmediatas.

Desarrollo de productos

Nuestros ingenieros se esfuerzan por resolver problemas del mundo real. Los derechos humanos guían nuestro trabajo de desarrollo de prácticas responsables de innovación, y estamos comprometidos con la diligencia debida en relación con los productos nuevos.

En 2023, lanzamos numerosos productos nuevos fascinantes, como [Threads](#) en julio, los [lentes inteligentes Ray-Ban Meta](#) y los [canales de WhatsApp](#) en septiembre y [Quest 3](#), nuestro visor de realidad mixta, en octubre. Las consideraciones de derechos humanos estuvieron integradas en el proceso de desarrollo de cada uno de estos productos.

Además, seguimos construyendo el metaverso, que incluye iniciativas para mejorar la educación y contrarrestar el sesgo y el odio. Por ejemplo, lanzamos varias experiencias de realidad virtual para crear comunidades, como [MLK: Now is the Time, Inside the Mosque](#) (una experiencia de [dos partes](#)) y [ABLE](#).

En un entorno rápido e iterativo, los equipos crean, prueban, mejoran y despliegan productos y servicios (incluidos aquellos que facilita la IA). Creamos desde una [perspectiva de privacidad](#) y diseñamos enfoques de mitigación del riesgo de los productos que coinciden con las necesidades concretas de nuestro proceso de desarrollo dinámico de productos. Al principio del proceso de desarrollo, los equipos de productos aplican recomendaciones sobre innovación responsable en línea con los principios de derechos humanos. Esto ayuda a que los

Riesgos significativos relevantes:

- Libertad de opinión y expresión
- Privacidad
- Igualdad y no discriminación



equipos puedan anticipar y mitigar daños potenciales a las personas, las comunidades y la sociedad. Nuestros expertos en derechos humanos y civiles pueden proporcionar orientación o realizar sprints rápidos, por ejemplo, para identificar los países de mayor riesgo para los lanzamientos de productos.

Por ejemplo, durante el proceso de desarrollo del producto Threads, los expertos en derechos humanos usaron los procesos de revisión de políticas y productos existentes para evaluar el impacto potencial en los derechos humanos. Trabajaron con el equipo de Threads en la implementación de mitigaciones, incluidas las relacionadas con solicitudes gubernamentales potenciales para censurar contenidos.

También seguimos desarrollando Project Height, un marco para que los equipos de productos evalúen las preocupaciones relacionadas con los derechos civiles que se presentan en los lanzamientos de productos. Esto les permite a los equipos de desarrollo de productos crear teniendo en cuenta los derechos civiles durante el proceso de desarrollo. Se complementa con otros procesos de riesgo que usamos para evaluar la seguridad de nuestros productos.



MLK: Now is the Time by Meta Quest



Riesgos significativos relevantes:

- Vida, libertad y seguridad de la persona
- Derecho a la participación pública, al voto y a presentarse a una candidatura
- Libertad de opinión y expresión
- Igualdad y no discriminación
- Privacidad
- Interés superior del niño

Cómo nos preparamos para las situaciones críticas y las crisis mundiales y cómo respondemos a ellas

Trabajamos en toda la empresa para incorporar los principios de derechos humanos en nuestras iniciativas de integridad³.

Esto incluye cómo nos preparamos para enfrentar desafíos mayores en países de alto riesgo, así como para acontecimientos críticos planificados, como las elecciones, y nuestra respuesta coordinada a eventos de emergencia. Este trabajo coordinado entre equipos especializados nos ayuda a anticiparnos y responder de forma eficaz a las tensiones crecientes dondequiera que se produzcan.

Cómo se prepara Meta para las elecciones

Muchos países celebraron elecciones en 2023 y muchos más se preparan para un intenso ciclo electoral en 2024. En 2023, Meta facilitó la preparación para más de 130 elecciones en todo el mundo, incluidos Pakistán, Argentina, Turquía y Nigeria. A continuación, compartimos información valiosa y las medidas que tomamos en relación con dos elecciones relevantes en Brasil y Pakistán.

Nuestros equipos especializados lideran los preparativos en Meta para implementar protecciones de cara a las elecciones de todo el mundo. Estas pueden incluir operaciones de seguridad avanzadas para enfrentar amenazas

adversas, políticas actualizadas y procesos para eliminar el contenido potencialmente dañino de nuestras plataformas, nuestra red mundial de verificación de datos líder del sector, y transparencia relacionada con los anuncios sobre temas políticos y sociales.

También evaluamos continuamente los riesgos de daños inminentes durante acontecimientos críticos, como las elecciones, para poder responder con medidas de productos y políticas específicas limitadas en el tiempo que mantengan la seguridad de las personas.

Si bien la mayor parte de nuestro trabajo es proactiva y se realiza con bastante antelación a las fechas de elecciones, también nos preparamos para reaccionar a situaciones de alto riesgo. En dichas situaciones, Meta tiene varios enfoques que podemos usar para abordar diferentes escenarios. Por ejemplo, cambiar o limitar funciones de los productos a través de límites en la frecuencia de los mensajes y en la distribución de contenidos.

Algunas de las maneras en que identificamos tendencias potencialmente dañinas consisten en recopilar información de informes públicos, revisar las recomendaciones de nuestros socios de confianza, observar continuamente las tendencias de contenido, aplicar la diligencia debida en materia de derechos humanos y evaluar los análisis de nuestros expertos en inteligencia. Esta información, junto con las Normas comunitarias y los sistemas de aplicación de políticas

³"Integridad" es un término técnico en Meta. Meta cuenta con una red de equipos que muchas veces incluyen la palabra "integridad" en sus nombres y que gestionan temas relacionados con la privacidad, la confianza y seguridad, la protección de cuentas y computadoras, la reducción de experiencias negativas y más. Entre muchas otras cosas, estos equipos crean herramientas para evitar daños, moderar las plataformas y hacer cumplir nuestras políticas.

existentes, ayuda a conformar el tipo de mitigaciones que podremos aplicar en políticas y productos para evitar el abuso durante elecciones de alto riesgo. Esto incluye nuestro [protocolo de la política de crisis](#) y medidas independientes que podemos tomar para adaptar los sistemas en respuesta al [mayor riesgo](#). También proporcionamos herramientas de transparencia, como la [biblioteca de anuncios](#).

Brasil

Empezamos a prepararnos para las [elecciones presidenciales de Brasil de 2022](#) con un año de anticipación. Mantuvimos las iniciativas durante el período posterior a las elecciones, lo que nos permitió responder rápidamente a los [ataques del 8 de enero de 2023](#) al Congreso, la Corte Suprema y otros edificios públicos. Meta también designó los disturbios posteriores a las elecciones como una crisis conforme al protocolo de la política de crisis para ayudar a la empresa a evaluar el modo de mitigar los riesgos relativos al contenido.


Los preparativos para las elecciones de 2022 incluyeron la elaboración de escenarios de riesgo durante las elecciones y con posterioridad a ellas. Consideramos las normas internacionales sobre libertad de expresión y otros derechos humanos. También realizamos cambios en los productos y adaptamos políticas para


proteger la integridad de las elecciones de Brasil. Esto incluyó asociarnos con el Tribunal Electoral Superior para proporcionar acceso a información confiable sobre la votación.

Desplegamos una variedad de herramientas y métodos para evitar el lenguaje que incita al odio, la incitación a la violencia y la desinformación potenciales antes de los ataques del 8 de enero de 2023, durante dichos ataques y después de ellos. Designamos a Brasil como un lugar de alto riesgo temporal poco antes de las elecciones, de forma que eliminamos incentivos a portar armamentos o ingresar por la fuerza a edificios gubernamentales, tal como señaló el [Consejo asesor de contenido en la decisión del caso "Discurso del general brasileño"](#).


Desde el inicio de las campañas electorales el 16 de agosto de 2022 hasta fines de enero de 2023, [eliminamos](#) más de 1 millón de contenidos en Facebook y más de 960.000 contenidos en Instagram por infringir las políticas de violencia e incitación en Brasil. Dichos contenidos incluían mensajes en los que se pedía la intervención militar.


Durante este período de cinco meses en Brasil, eliminamos contenidos que infringían las políticas sobre lenguaje que incita al odio:

 +570.000

 +520.000

También eliminamos contenidos en Brasil por infringir nuestras políticas de bullying y acoso:

 +380.000

 +630.000

Violencia de género

La amenaza de acoso y violencia de género contra las candidatas, periodistas y defensoras de derechos humanos muchas veces aumenta durante las elecciones. Para mitigar estos riesgos y como parte de un enfoque más amplio para incrementar la protección de las mujeres en las plataformas y conectarlas con los servicios públicos, las medidas que tomamos en Brasil incluyeron las siguientes:

- Asociarnos con el Ministerio de las Mujeres para lanzar [un canal oficial de WhatsApp en Brasil](#) para establecer un medio de comunicación adicional con la Defensoría Pública en el que las mujeres puedan presentar denuncias o solicitar información sobre la violencia contra la mujer. Las usuarias pueden obtener información sobre las leyes relevantes, encontrar direcciones de servicios dedicados a la asistencia de mujeres y hablar directamente con auxiliares.
- Publicar [una guía online](#) en portugués para proteger a las mujeres contra la violencia en internet.
- Trabajar con grupos del sector, organizaciones de la sociedad civil y autoridades reguladoras, incluido el Ministerio de las Mujeres, en iniciativas conjuntas con el objetivo de abordar el acoso y la violencia de género.



Pakistán

Aunque las elecciones de Pakistán se realizaron en 2024, nuestros preparativos electorales empezaron en 2022. Estos incluyeron iniciativas en toda la empresa en productos, políticas y operaciones para evitar y mitigar los riesgos para los derechos humanos relacionados con el uso de nuestras plataformas.

Entre otras medidas, desarrollamos señales para identificar el contenido con cuestiones relacionadas con los derechos y escalamientos de alta prioridad. De conformidad con los [principios de la GNI](#), publicamos un [caso práctico](#) en tiempo real en nuestro Centro de transparencia sobre las solicitudes de eliminación enviadas en diciembre de 2023 por el Gobierno de Pakistán de contenidos que no infringían nuestras Normas comunitarias ni la legislación local.

La participación de las partes interesadas fue fundamental. Antes de las elecciones, organizamos reuniones informativas con las partes interesadas de la región en materia de derechos humanos sobre las iniciativas de integridad de las elecciones. También trabajamos con la Comisión de Elecciones para explicar nuestro enfoque de las solicitudes gubernamentales de medidas y los derechos humanos. Se limitó el acceso a internet durante el período previo a las elecciones, y Meta compartió [información](#) con las partes interesadas sobre cómo configurar un servidor proxy para conectarse a WhatsApp.

Como miembro de grupos de varias partes interesadas, como la [GNI](#) y la [Asia Internet Coalition](#), Meta participó en declaraciones de preocupación sobre los impactos en los derechos a causa de interrupciones en internet y legislación que infringe los derechos antes de las elecciones.

Varios meses antes de las elecciones, también tomamos medidas para proporcionar a los defensores de derechos humanos y otras personas vulnerables mayores protecciones en la plataforma. Esto incluyó la protección contra la eliminación de contenido basada en acusaciones falsas de infracción de las Normas comunitarias. Nuestro enfoque fue inclusivo, conforme a nuestra [Política corporativa de derechos humanos](#), que incorpora la definición amplia de defensores de los derechos humanos de la [Declaración sobre los Defensores de los Derechos Humanos de las Naciones Unidas](#).



Preparación para los desafíos en países que necesitan una respuesta activa

La clasificación de países que necesitan una respuesta activa se determina a través de un proceso basado en evidencia, dirigido por equipos especializados y nuestros expertos en derechos humanos, y se conforma a través del [proceso de diligencia debida en materia de derechos humanos](#) para que se puedan revisar y priorizar los países con el mayor riesgo de violencia y daño en la vida real. Esto activa una estrategia de mitigación de riesgos en profundidad. Puede incluir la supervisión intensificada de riesgos e iniciativas de mitigación, además de inversiones adicionales, como en la revisión de contenido en idiomas relevantes para el conflicto a fin de eliminar contenido que infringe las políticas y brindar ayuda de productos personalizada. Dada la complejidad de los riesgos, también recurrimos a socios de confianza y verificadores de datos externos para que colaboren con nuestras iniciativas en los países que necesitan una respuesta activa.



Respuesta a las crisis mundiales

Respondemos a crisis inesperadas en todo el mundo, incluidos los conflictos interestatales, los atentados terroristas y las crisis no violentas como aquellas relacionadas con desastres ambientales. Nuestras iniciativas de respuesta a crisis involucran a expertos en áreas como política de contenido, operaciones internacionales, derechos humanos y producto, así como otros expertos internos que evalúan y responden a las situaciones según el [protocolo de la política de crisis](#). Como se describe en nuestro [informe sobre derechos humanos de 2022](#), el protocolo guía nuestro uso ágil de medidas normativas excepcionales o específicas para mitigar el daño potencial según los riesgos observados y las intervenciones en situaciones de crisis anteriores y principios de derechos humanos.

Nuestra evaluación integral de riesgos significativos para los derechos humanos nos ayuda a priorizar entre diferentes riesgos potenciales para los derechos humanos durante situaciones de conflicto, por ejemplo, entre el derecho a la libertad y la seguridad de las personas, el derecho a la libertad de expresión, y la seguridad y el bienestar de nuestras comunidades. Como parte de nuestro conjunto de herramientas, podemos adoptar una serie de medidas temporales para ayudar a mantener la seguridad de las personas y mitigar el riesgo de que nuestras plataformas puedan utilizarse para exacerbar aún más las tensiones online y offline. Podemos considerar acontecimientos relevantes del derecho internacional público, incluido el derecho internacional humanitario (el derecho de los conflictos armados) de conformidad con la guía que proporcionan los Principios

Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas acerca de una mayor diligencia debida en materia de derechos humanos en épocas de conflicto. A continuación exponemos ejemplos de nuestro trabajo para fines ilustrativos. En la subsección



siguiente, proporcionamos [actualizaciones](#) sobre nuestras medidas de diligencia debida en materia de derechos humanos en Israel y Palestina y el conflicto relacionado.

Sudán

En abril de 2023, estallaron enfrentamientos entre las Fuerzas Armadas de Sudán (SAF) y las Fuerzas de Apoyo Rápido (RSF), lo que causó desplazamientos masivos, inseguridad alimentaria, falta de acceso a la atención médica y otros desafíos que requieren ayuda humanitaria. Utilizando el protocolo de la política de crisis de Meta, se clasificó a Sudán en el nivel más alto de riesgo.

Riesgos significativos relevantes:

- Vida, libertad y seguridad de la persona
- Libertad de opinión y expresión
- Libertad de reunión y asociación
- Privacidad
- Derecho a la salud



Con el protocolo de la política de crisis como guía, tomamos medidas para que se respeten los derechos humanos y para minimizar los riesgos de daño, así como también para colaborar en la provisión de ayuda humanitaria.

En agosto de 2023, designamos a las Fuerzas de Apoyo Rápido conforme a nuestra [política de personas y organizaciones peligrosas](#), que no permite la presencia en las plataformas de Meta de organizaciones o personas que reivindicuen una misión violenta o cometan actos de violencia. Esto ayudó a combatir la propagación de contenido potencialmente dañino. La decisión se tomó después de trabajar con agencias de las Naciones Unidas y sobre la base de la investigación y los informes de diligencia debida en materia de derechos humanos por parte de organizaciones no gubernamentales sobre las infracciones graves de los derechos humanos. El Consejo asesor de contenido revisó y avaló la decisión. También designamos a Sudán como lugar de alto riesgo temporal, lo que ayudó a eliminar contenido que incitaba a los civiles a armarse o a llevar armas a un lugar determinado en el país. Trabajamos con verificadores de datos externos para desmentir, etiquetar y reducir la visibilidad de la información errónea relacionada con el conflicto en la plataforma. Además, eliminamos la información errónea que, según los socios de confianza, podía contribuir al riesgo de daño físico o violencia inminente.

Debido al desplazamiento masivo desde la capital, Jartum, resultó difícil o imposible contactar a muchas de nuestras partes interesadas, incluidos los funcionarios del Gobierno, la sociedad civil, los medios de comunicación y los defensores de derechos humanos. Esto nos obligó a identificar y trabajar con grupos nuevos para apoyar las iniciativas humanitarias, así como también ayudar a identificar y detener la difusión de contenido potencialmente dañino y la información errónea. Trabajamos con los grupos de sociedad civil más activos, incluidos médicos, medios de comunicación,

ciudadanos periodistas y grupos de mujeres, tanto en Sudán como principalmente en la diáspora. Les informamos sobre la política de contenido y les proporcionamos capacitaciones sobre seguridad digital y las herramientas para crisis de Meta. Para ello, organizamos talleres sobre el uso de las Comunidades de WhatsApp, que luego se utilizaron para coordinar las evacuaciones y la ayuda humanitaria.

A pesar del deterioro de la seguridad, nuestros socios de confianza continuaron reportando contenido relacionado con la violencia e incitación, información errónea, lenguaje que incita al odio, y bullying y acoso. Las evaluaciones de los socios de confianza ayudaron a aplicar la política de información errónea y daños al contenido que utilizaba imágenes fuera de contexto para avivar las tensiones, así como también a las afirmaciones hechas en contra de las agencias humanitarias que operaban en Sudán.

En consonancia con nuestro enfoque sobre [otros conflictos](#), eliminamos parte del contenido gráfico infractor sin aplicar [faltas](#) (penalizaciones a las infracciones que al acumularse generan mayores [restricciones a las cuentas](#)) para evitar penalizar o restringir de forma excesiva a los usuarios que intentaban crear conciencia sobre los impactos del conflicto. Una recomendación del Consejo asesor de contenido en el [caso del video gráfico de Sudán](#) en 2022 dio forma a este enfoque. Este caso motivó una revisión en 2023 de nuestras políticas sobre el uso compartido de contenido violento y gráfico para crear conciencia sobre los abusos de los derechos humanos, junto con una explicación más detallada acerca de cómo aplicamos nuestras [excepciones por interés periodístico](#).

Debido a la grave escasez de medicamentos en el país causada por la falta de infraestructura de salud, hicimos cambios temporales en nuestras políticas para que los usuarios pudieran solicitar o donar medicamentos, lo cual generalmente está prohibido

Cómo desarrollar resiliencia: capacitación sobre seguridad digital y política de contenido

2023:

6

sesiones de capacitación realizadas

300

beneficiarios alcanzados

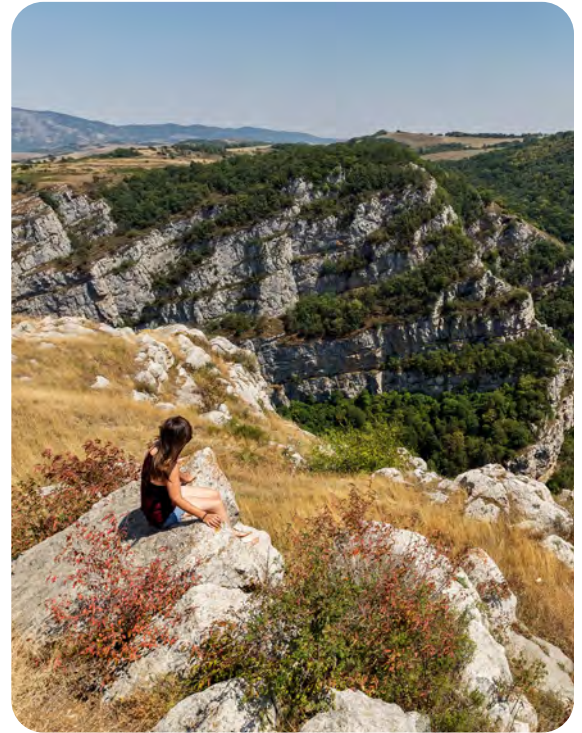
Después del conflicto:

2

sesiones de capacitación realizadas

70

beneficiarios alcanzados



por nuestras políticas de bienes restringidos. Además, trabajamos con el Sudan Medical Specialization Board para lanzar [SMSB Sudan Clinic](#), una app de telemedicina que utiliza la API de WhatsApp. Esta app se ofreció a través de nuestro programa de exención de tarifas por impacto social en WhatsApp, un programa destinado a socios gubernamentales y organizaciones sin fines de lucro mediante el cual eximimos las tarifas por el envío de mensajes.

También necesitábamos equilibrar los mensajes sobre movilidad y seguridad (la comunicación de información que ayuda a las personas a viajar mientras se garantiza su seguridad) sin apoyar la migración irregular hacia los países vecinos como la República Centroafricana, Chad, Egipto, Etiopía, Libia y Sudán del Sur. Esto se debe a que el riesgo de explotación de personas crece a medida que aumentan las necesidades de la población desplazada, lo que provoca la aparición de agentes malintencionados en estas zonas.

Nos esforzamos por identificar a los agentes malintencionados y eliminar contenidos que pudieran poner en peligro a las personas, incluidos los de tráfico ilegal de personas.

Nagorno Karabaj

Entre el 19 y el 20 de septiembre de 2023, Azerbaiyán inició una ofensiva militar en la autodeclarada República de Artsaj, parte de la región de Nagorno Karabaj en disputa. Dicha ofensiva provocó el éxodo de personas de la etnia armenia de la región, lo que generó una crisis de refugiados.

Un equipo interfuncional, que incluía expertos en derechos humanos, estuvo monitoreando la situación, movilizó rápidamente los recursos y aplicó nuestro [protocolo de la política de crisis](#). Expertos internos con habilidades lingüísticas y conocimiento contextual gestionaron el aumento de la prevalencia del lenguaje que incita al odio, el bullying, el contenido gráfico y las referencias a personas

Riesgos significativos relevantes:

- Libertad de opinión y expresión
- Vida, libertad y seguridad de la persona
- Privacidad



y organizaciones peligrosas. También colaboramos con agencias de las Naciones Unidas presentes en la zona para entender más a fondo la dinámica de la crisis. Meta proporcionó protecciones adicionales para los actores cívicos de ambos bandos a fin de reducir los riesgos de suplantación de identidad, acoso y hackeos.

Como en otros conflictos, tuvimos que decidir el modo de equilibrar la protección de la privacidad y la dignidad de los prisioneros de guerra con los beneficios de documentar abusos de los derechos humanos. Las Normas comunitarias de Meta prohíben publicar información que podría revelar las identidades o ubicaciones de prisioneros de guerra. Esto está en consonancia con las responsabilidades de la empresa en materia de derechos humanos en virtud de los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas. Estas responsabilidades se acentúan durante los conflictos armados y deben regirse por las normas del derecho internacional humanitario. Revelar las identidades y ubicaciones de prisioneros de guerra puede poner en riesgo su seguridad, privacidad y dignidad, y la seguridad de sus familias.

Al mismo tiempo, dicha exposición puede dar forma al debate público y generar conciencia sobre posibles malos tratos, incluidas las infracciones de derechos humanos internacionales y el derecho internacional humanitario. También puede impulsar la adopción de medidas que protejan los derechos y garanticen la responsabilidad. La escala y la rapidez con la que se pueden compartir imágenes de prisioneros de guerra a través de los medios sociales complican la tarea de resolver estos intereses en conflicto.

En 2023, el [Consejo asesor de contenido](#) mantuvo la decisión de conservar una publicación de Facebook con un video que mostraba a prisioneros de guerra identificables con la pantalla de advertencia que indica que es "contenido perturbador". El Consejo coincidió con Meta y señaló que, en esta situación, las publicaciones podían nutrir el debate público y generar conciencia sobre posibles malos tratos, incluidas infracciones de los derechos humanos internacionales y del derecho internacional humanitario, lo cual superaba el riesgo para la seguridad y dignidad de los prisioneros de guerra.

Retención de evidencia

Brindamos apoyo a la justicia en todos los delitos internacionales. Llevamos varios años explorando iniciativas que respeten los derechos para la retención y divulgación de evidencia, en consulta exhaustiva con la sociedad civil, el mundo académico y expertos y organismos internacionales en materia penal.

Como [mencionamos](#) en la respuesta a la recomendación del Consejo asesor de contenido, nos esforzamos en crear un enfoque que facilite la labor de los tribunales internacionales y los mecanismos de responsabilidad autorizados por las Naciones Unidas (como misiones de determinación de hechos y comisiones de investigación) para que nos soliciten la retención extendida de datos relevantes para sus investigaciones en curso. Este trabajo está en gran parte completado y es distinto de nuestras políticas [de larga data para responder a solicitudes de conservación por parte de las autoridades](#).

Informamos a una variedad de mecanismos autorizados y relatores especiales de las Naciones Unidas sobre nuestro enfoque y describimos el proceso para realizar solicitudes de conservación prolongada a Meta. Revisaremos con cuidado todas las solicitudes recibidas para asegurarnos de que se ajusten a nuestras políticas y a la legislación aplicable.

Se trata de un área nueva sin prácticas recomendadas establecidas o probadas, y aún existen consideraciones legales, de privacidad y de políticas significativas inherentes a este trabajo. Compartiremos más novedades sobre el trabajo en esta área en las [actualizaciones semestrales sobre el Consejo asesor de contenido](#) y en nuestro informe anual sobre derechos humanos.



Riesgos significativos relevantes:

- Vida, libertad y seguridad de la persona
- Privacidad

Migración en América Central y América del Sur

El tráfico ilegal de personas puede generar la explotación de aquellas personas vulnerables que intentan abandonar su país de origen, con frecuencia en búsqueda de una mejor calidad de vida. Nos esforzamos por eliminar contenido que facilite o coordine la explotación de personas, lo que incluye el tráfico ilegal y la trata de personas.

En 2023, identificamos riesgos potenciales para los derechos humanos en relación con el tráfico ilegal de personas migrantes a lo largo de América del Norte, América Central y América del Sur, incluida la región del Darién, un tramo de selva densa y peligrosa que se extiende por el norte de Colombia y el sur de Panamá. Se trata de zonas de mayor riesgo de actividad comercial ilegal, presencia online de organizaciones peligrosas e información errónea para migrantes y refugiados, especialmente cuando algunos intentan llegar a la frontera entre Estados Unidos y México.

En respuesta a dichos riesgos, aplicamos nuestras [Normas comunitarias](#) y políticas para empresas sobre [comercio](#) que prohíben la explotación de personas, incluido el tráfico ilegal de personas. Colaboramos con la sociedad civil y las autoridades para hacer frente a la información errónea y seguimos supervisando nuestras políticas sobre contenidos orgánicos, pagados y comerciales que infrinjan las normas. También facilitamos que verificadores de datos externos busquen y califiquen contenido relacionado con la frontera entre Estados Unidos y México, ya que reconocemos la importancia de la celeridad durante situaciones críticas. Utilizamos la detección de palabras clave para agrupar el contenido relacionado en un solo lugar para nuestros socios de verificación de datos, que clasificaron contenidos en nuestra plataforma y publicaron artículos en inglés y español acerca de la información errónea.

También apoyamos a las comunidades de migrantes y refugiados y las organizaciones



no gubernamentales relacionadas a través de programas educativos y de integridad como Soy Digital, que se enfoca en las competencias digitales para identificar información errónea, fraudes y estafas.

Evaluaciones sobre el impacto en los derechos humanos

Las metodologías de diligencia debida en materia de derechos humanos están evolucionando. Consideramos los riesgos potenciales y aprendemos de iniciativas anteriores para decidir cómo mejorar nuestro enfoque y afianzar las iniciativas que dieron buenos resultados.

Realizamos evaluaciones sobre el impacto en los derechos humanos en países específicos, apps, servicios, hardware o iniciativas estratégicas, según lo establecido por nuestra evaluación integral de riesgos significativos para los derechos humanos. Dichas evaluaciones nos ayudan a anticipar impactos posibles, en especial cuando preparamos funciones y productos nuevos, como el [cifrado de extremo a extremo](#). Las evaluaciones sobre el impacto en los derechos humanos involucran la [participación de partes interesadas](#), un aspecto fundamental de la diligencia debida en materia de derechos humanos. Cuando realizamos una evaluación, cada recomendación suele requerir varios equipos y flujos de trabajo para implementarla en toda la empresa.

Compartimos resúmenes y recomendaciones de evaluaciones sobre el impacto en los derechos humanos anteriores que figuran en el [anexo](#). También proporcionamos actualizaciones sobre las medidas tomadas en respuesta a las evaluaciones sobre riesgos

potenciales para los derechos humanos relacionados con las plataformas de Meta en [Filipinas](#), [India](#), el [cifrado de extremo a extremo](#), [Israel y Palestina](#).

Israel y Palestina: actualización sobre nuestras medidas

Poco después de la publicación de nuestra [actualización de diligencia debida en materia de derechos humanos de Israel y Palestina del 2023](#), Hamás llevó a cabo los atentados terroristas del 7 de octubre contra Israel, seguidos por la respuesta militar de Israel en Gaza y una mayor participación y escalada por parte de otros agentes regionales. Sabemos que las personas de la región y alrededor del mundo se sintieron profundamente impactadas por nuestra respuesta a la situación de conflicto en curso en dicha región.

Inmediatamente después del atentado, Meta designó la violencia al nivel más alto de nuestro protocolo de la política de crisis e implementó medidas de respuesta a crisis inmediata, lo que incluyó un equipo interfuncional dedicado las 24 horas del día, todos los días. Los principios fundamentales de derechos humanos guiaron nuestras acciones. Usamos los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#) como base de nuestra [Política corporativa de derechos humanos](#) para priorizar y mitigar los riesgos más significativos para los derechos humanos. El derecho internacional humanitario también fue una referencia importante.

En un principio, compartimos los detalles sobre nuestra respuesta en una publicación de blog en [inglés](#), [árabe](#) y [hebreo](#) el 13 de octubre

Riesgos significativos relevantes:

- Vida, libertad y seguridad de la persona
- Libertad de opinión y expresión
- Igualdad y no discriminación
- Privacidad



de 2023 y proporcionamos más actualizaciones el 18 de octubre, el 5 de diciembre y el 8 de diciembre. Seguimos mejorando nuestro enfoque para reflejar las dinámicas cambiantes, incluyendo la crisis humanitaria en curso en Gaza y los rehenes aún retenidos por Hamás.

Mantener el equilibrio entre la seguridad de nuestras comunidades y la libertad de expresión en nuestras plataformas es particularmente complicado en situaciones de conflicto, y más aún cuando involucra a entidades sancionadas por EE. UU., como Hamás, Hezbolá y la Yihad Islámica Palestina, que también están designadas como organizaciones peligrosas según nuestras políticas.

El trabajo que ya llevamos a cabo en respuesta a la [diligencia debida en materia de derechos humanos de Israel y Palestina](#) ayudó a darle forma a nuestro enfoque. Por ejemplo, logramos redirigir el contenido en árabe de manera más adecuada a través de nuestros sistemas para conseguir una mayor precisión. También aumentamos los recursos de moderación de contenido en el mercado hebreo. Para obtener más detalles sobre las actualizaciones, consulta la [Actualización de Meta de 2024: diligencia debida en materia de derechos humanos de Israel y Palestina](#).

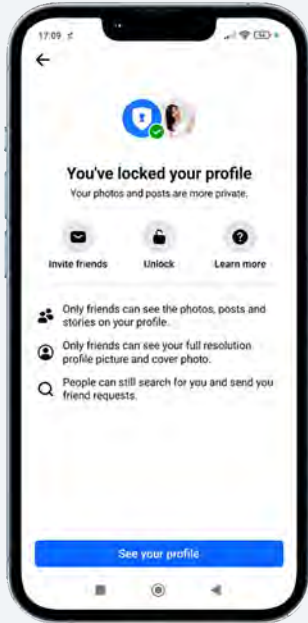
Nuestra respuesta al conflicto superó las medidas previstas en la [diligencia debida](#) y estuvo igualmente orientada por el [protocolo de la política de crisis](#). Adoptamos varias [medidas temporales de productos y políticas](#) para mitigar los riesgos significativos para los derechos humanos. Sabemos que en algunos casos estas medidas (por ejemplo, bajar los umbrales para la aplicación automática de políticas) pueden limitar involuntariamente el debate de situaciones mundiales críticas, lo que afecta la expresión de opiniones. Por este motivo, buscamos que nuestras medidas sean de duración limitada y proporcionales a los riesgos que conocemos.

Medidas temporales de políticas

Inicialmente nos basamos en la política de personas y organizaciones peligrosas y ampliamos la política de violencia e incitación para eliminar contenido que mostraba a rehenes identificables siendo secuestrados o retenidos en cautiverio, incluso cuando se hacía para condenar dichos hechos o generar conciencia sobre la situación. Lo hicimos para proteger la privacidad y la dignidad de las víctimas conforme a las normas del derecho internacional humanitario y para evitar que la propaganda de Hamás apareciera en nuestras plataformas, de acuerdo con la política de personas y organizaciones peligrosas.

A medida que la guerra continuaba, observamos que las personas comenzaron a compartir contenido sobre rehenes para refutar narrativas emergentes que negaban los ataques terroristas del 7 de octubre. En consecuencia, adoptamos un enfoque más matizado y permitimos que las personas compartieran contenidos que mostraban el secuestro de rehenes a fin de generar conciencia sobre los atentados o condenarlos. Seguimos eliminando contenido y priorizando la seguridad cuando la intención del usuario al compartirlo no era clara.

Realizamos otros cambios con el fin de proteger la expresión. Para algunas áreas de nuestras políticas, como ciertos tipos de [contenido gráfico y violento](#), eliminamos el contenido infractor sin aplicar [faltas](#) (penalizaciones a las infracciones que al acumularse generan mayores [restricciones a las cuentas](#)) para evitar penalizar o restringir de forma excesiva a los usuarios que intentaban generar conciencia sobre los impactos del conflicto.



Medidas temporales de productos

Adoptamos varias medidas temporales de productos para mantener a las personas seguras y mitigar riesgos significativos para los derechos humanos, como se detalla en nuestra publicación del [blog que describe nuestra respuesta al conflicto](#). Esto incluyó ajustes en los umbrales de confianza para la toma de acciones automáticas sobre el contenido, el bloqueo de ciertos hashtags en las búsquedas y cambios en los productos para abordar comentarios no deseados y problemáticos, además de medidas de seguridad como la herramienta [Restringir tu perfil](#).

También bajamos temporalmente el umbral por el que el contenido límite o potencialmente infractor (como imágenes o video que muestran violencia gráfica) podría ser considerado no apto para ser recomendado. Esta medida se aplicó al contenido no conectado, es decir, contenido de personas a quienes el usuario no sigue, pero que podrían aparecer en entornos como el feed, la búsqueda, la sección "Explorar" o Reels.

Al igual que con nuestras medidas temporales de políticas, introdujimos otros cambios destinados a proteger la expresión. En respuesta a un gran aumento en el uso de nuestros productos, ajustamos temporalmente algunos límites automatizados de frecuencia diseñados para prevenir comportamientos definidos como [spam en nuestras políticas](#) (como la frecuencia alta de publicación) para hacerlos más permisivos y reducir el riesgo de restricciones a usuarios legítimos.

Por otro lado, nuestras políticas generales sobre [recomendabilidad](#) y [degradación](#), que *no* forman parte de nuestras acciones de respuesta a crisis, pueden afectar la visibilidad de ciertos contenidos relacionados con el conflicto, como contenido gráfico y violento.

Meta llevó a cabo investigaciones frecuentes para evaluar el rendimiento de los sistemas de moderación de contenido y si estaban aplicando las políticas al contenido de forma excesiva o insuficiente. De cara al futuro, identificamos áreas para seguir mejorando, como priorizar la revisión de contenido en árabe según el dialecto y mejorar la supervisión de nuestros bancos de imágenes y videos para mitigar los riesgos potenciales de una aplicación excesiva de las políticas y los impactos relacionados en la libertad de expresión. Seguimos trabajando en ello.

Para obtener información más detallada sobre la respuesta, consulta nuestras cartas de respuesta a las preguntas que recibimos por parte de organizaciones de la sociedad civil [aquí](#) y [aquí](#). Para obtener más detalles sobre nuestras novedades de diligencia debida en materia de derechos humanos de Israel y Palestina de 2024, haz clic [aquí](#).



Riesgos significativos relevantes:

- Privacidad
- Libertad de opinión y expresión
- Libertad de reunión y asociación
- Igualdad y no discriminación

Cifrado de extremo a extremo: actualización sobre nuestras medidas

En diciembre de 2023, comenzamos a implementar el cifrado de extremo a extremo de forma predeterminada en todos los chats y llamadas personales de Facebook y Messenger. Esto marcó un hito importante en el recorrido que comenzamos en 2019 para expandir el cifrado de extremo a extremo de forma predeterminada a todas nuestras apps de mensajes. También seguimos desarrollando el cifrado de extremo a extremo para Instagram Direct. WhatsApp proporciona cifrado de extremo a extremo de forma predeterminada desde 2016.

El cifrado de extremo a extremo mejora la seguridad y la privacidad, ya que asegura que solo tú y el destinatario puedan ver tus mensajes. Esto brinda una capa adicional de seguridad que protege los mensajes que intercambias con familiares y amigos, y significa que nadie, ni siquiera Meta, puede ver o escuchar lo que se envía o se dice.

Cuando reportas una conversación cifrada de extremo a extremo, los mensajes recientes de esa conversación se envían de forma segura desde tu dispositivo a nuestros equipos de ayuda.

El marco de derechos humanos fue clave para orientar la expansión del cifrado de extremo a extremo. El cifrado de extremo a extremo permite directamente el derecho a la privacidad que, a su vez, permite que se materialice una amplia variedad de otros derechos humanos, lo que incluye la libertad de expresión, de asociación, de opinión, de movimiento y la seguridad física. Al mismo tiempo, siempre existe el riesgo de que alguien abuse de cualquier servicio de mensajería, incluidas las funciones cifradas, para dañar a otros.

En 2022, compartimos los resultados completos de una evaluación independiente sobre el impacto en los derechos humanos realizada por BSR, que solicitamos para analizar los beneficios y riesgos potenciales para los derechos humanos en relación con nuestros planes de ampliar el cifrado de extremo a extremo predeterminado. De la evaluación surgieron 45 recomendaciones para la maximización de los beneficios potenciales y la mitigación de impactos adversos potenciales por parte de Meta. Desde nuestra respuesta inicial a estas recomendaciones, seguimos trabajando para implementar los compromisos que asumimos y proporcionamos actualizaciones públicas exhaustivas, entre ellas, en enero y agosto de 2023, así como también un informe de seguridad detallado.



Temas destacados

Riesgos significativos relevantes:

- Privacidad
- Vida, libertad y seguridad de la persona

Combatir las amenazas de seguridad

En los últimos seis años, Meta dio a conocer informes públicos sobre iniciativas para detectar y combatir amenazas a la seguridad en nuestras plataformas. El análisis y las investigaciones posteriores a la interrupción formaron parte de la base para que Citizen Lab y 45 grupos de la sociedad civil presentaran un caso en la Cumbre por la Democracia para prohibir la venta de spyware hasta que se implementen medidas de seguridad que eviten los abusos de los derechos humanos. Luego, el presidente de los EE. UU. Joe Biden dictó un decreto para prohibir a las entidades federales de EE. UU. el uso de spyware que permitiera abusos de los derechos humanos.

Algunos ejemplos de las amenazas que mitigamos en 2023 se relacionan con el sector de vigilancia por contratación. Las empresas de vigilancia por contratación incluyen aquellas que ofrecen servicios para supervisar y rastrear a organizaciones o personas con diversos fines. Existen preocupaciones respecto de la privacidad y el uso indebido de la vigilancia por contratación por parte de Gobiernos autoritarios, organizaciones delictivas y otros agentes malintencionados que intentan espiar a quienes son críticos del régimen, figuras de la oposición, periodistas y

defensores de los derechos humanos con el fin de restringir la libertad de expresión.

En 2023, identificamos y eliminamos seis redes de cuentas en Facebook e Instagram vinculadas con ocho empresas del sector de vigilancia por contratación de Italia, España y los Emiratos Árabes Unidos. Entre sus tácticas se encontraban la ingeniería social y el phishing en internet usando perfiles que se hacían pasar por manifestantes, periodistas y mujeres jóvenes para engañar a las personas con el fin de que compartieran sus correos electrónicos y números de teléfono, o que hicieran clic en enlaces maliciosos.

"El crecimiento de la industria del spyware privado facilita el acceso a funciones sofisticadas de vigilancia mientras oculta a los clientes que compran y despliegan esos servicios. Este tipo de vigilancia indiscriminada y poco transparente representa una amenaza significativa para los defensores de los derechos humanos. Por eso, nos comprometemos a trabajar con socios de la sociedad civil y del sector para exigirles a las empresas de spyware abusivo que se asuman sus responsabilidades".

David Agranovich, director, Threat Disruption, Meta



Riesgos significativos relevantes:

- Interés superior del niño
- Privacidad
- Libertad de opinión y expresión

Seguridad infantil y juvenil

La seguridad online para los niños y jóvenes⁴ es un desafío en todo internet que requiere el esfuerzo y la colaboración entre el sector, el Gobierno y la sociedad civil, ya que buscamos ayudar a mantener seguros a los usuarios y al mismo tiempo proporcionar espacio para ejercer una variedad de derechos humanos, como la libertad de expresión y el acceso a la información, entre otros.

Llevamos una década trabajando en este tema y seguimos utilizando un enfoque basado en los derechos para el diseño de la política de contenido y los servicios. El [marco del Interés superior del niño](#) que utilizamos se basa en la [Convención sobre los Derechos del Niño de las Naciones Unidas](#), así como también el Código de Diseño Apropiado para la Edad del Reino Unido, los Principios Fundamentales para Niños de la Comisión de Protección de Datos de Irlanda y la recomendación sobre menores de la *Commission Nationale de l'Informatique et des Libertés* (CNIL) de Francia.

Explotación infantil

La seguridad infantil siempre es una prioridad para nosotros. Buscamos evitar los daños por medio de la aplicación de nuestras políticas y el desarrollo de herramientas de prevención. Trabajamos con diligencia para hacer que las herramientas para reportar explotación infantil sean más fáciles de encontrar, lo que llevó a un aumento del 75% en los reportes de adolescentes en el primer trimestre de 2023 en comparación con

⁴No existe una definición universal para la palabra "juven". La [Convención sobre los Derechos del Niño de las Naciones Unidas](#) define al "niño" como todo ser humano menor de 18 años de edad (artículo 1), lo que incluye a los adolescentes de más de 13 años. En este informe utilizamos indistintamente las palabras "juven" y "adolescente".

el mismo trimestre de 2022 en Messenger e Instagram. También implementamos formas innovadoras adicionales para buscar, eliminar o restringir proactivamente cuentas que podrían infringir nuestras políticas de seguridad infantil. Por ejemplo, lanzamos una tecnología proactiva para detectar y desactivar automáticamente cuentas con actividades e interacciones en la plataforma que demuestren interés sexual en niños, así como también las cuentas vinculadas a estas. Se identificaron y [eliminaron](#) automáticamente más de 90.000 cuentas desde el 1 de agosto de 2023 hasta el 31 de diciembre de 2023 gracias a este método. Nos esforzamos por detectar y restringir proactivamente cuentas de adultos ante los primeros indicios de interacciones no deseadas con adolescentes o sobre ellos, o intentos de entablar contactos potencialmente inseguros.





Como miembros fundadores de [Take It Down](#), una herramienta diseñada para la prevención proactiva de la difusión online de imágenes íntimas de personas jóvenes, trabajamos con el [Centro Nacional para Niños Desaparecidos y Explotados](#) para traducirla a más idiomas e incentivamos a más agentes del sector a unirse. De esta manera, logramos el acceso de millones de adolescentes más.

También somos miembros fundadores de [Lantern](#), un programa nuevo de la Tech Coalition que se lanzó en noviembre de 2023 y que les permite a las empresas de tecnología compartir indicios de cuentas y comportamientos abusivos en las plataformas online. Business for Social Responsibility realizó una [evaluación sobre el impacto en los derechos humanos](#) de este programa.

"Uno de los desafíos actuales más importantes de nuestro sector es evitar la explotación infantil. Los depredadores online son delincuentes decididos que utilizan diferentes apps y sitios web para llegar a las personas jóvenes. También ponen a prueba las defensas de cada plataforma y aprenden a adaptarse rápidamente. Por eso, ahora y como siempre, estamos trabajando arduamente para mantenernos a la vanguardia. Además de desarrollar tecnología que detecta a los depredadores, contratamos especialistas dedicados a la seguridad infantil online y compartimos información con nuestros colegas de la industria y las autoridades".

Antigone Davis, directora global de Seguridad, Meta



Seguridad y bienestar juvenil

Queremos que las personas jóvenes tengan una experiencia online segura y apropiada para su edad a la vez que se respetan sus derechos digitales. Nuestro enfoque se centra en primer lugar en evitar el daño, brindándoles a las personas herramientas para controlar la experiencia y respondiendo a las infracciones a las políticas con rapidez. Desarrollamos muchas [herramientas, funciones y recursos](#) para ayudar a adolescentes y padres. Seguimos trabajando en ello.

En 2023, seguimos ampliando nuestro trabajo en torno a las [herramientas de supervisión parental](#) en todos nuestros servicios. Realizamos [sesiones de diseño conjunto](#) para padres, adolescentes, tutores y expertos a través de [Trust, Transparency and Control Labs](#) (TTC Labs) (una iniciativa intersectorial para darles a las personas el control de su privacidad) y trabajamos con asesores juveniles para entender el comportamiento de los adolescentes y guiar el diseño de productos. También lanzamos las [cuentas administradas por los padres](#) para preadolescentes de 10 a 12 años en Meta Quest 2 y 3.

Consultamos con regularidad a expertos en desarrollo, psicología y salud mental adolescente para que nuestras plataformas se mantengan seguras y apropiadas para la edad de las personas jóvenes, lo que incluye el desarrollo continuo de nuestra comprensión sobre qué tipos de contenido pueden ser más adecuados para adolescentes. Nos reunimos con [socios de seguridad](#), grupos de jóvenes y asesores de todo el mundo, así como también con UNICEF para informarlos sobre nuestro trabajo e invitarlos a colaborar.

Teniendo en cuenta los puntos de vista de los expertos en cuentas, desarrollamos [protecciones nuevas](#) para que la experiencia en nuestras apps sea más apropiada para la edad, lo que incluye, por ejemplo, ocultar más resultados de búsquedas, indicar a los adolescentes que actualicen su configuración de privacidad en Instagram y colocar automáticamente a los adolescentes en la configuración de control de contenido más restrictiva en Facebook e Instagram.

Participación de partes interesadas

Realizamos [sesiones de diseño conjunto](#) a través de [TTC Labs](#) para invitar a los usuarios, la comunidad y los expertos a participar en la creación de experiencias online positivas para las personas jóvenes escuchando sus necesidades y preocupaciones para darle forma al diseño de productos.

87	adolescentes
71	tutores
221	expertos
13	países

Participación de partes interesadas



El trabajo proactivo con un amplio espectro de partes interesadas es fundamental para el enfoque de derechos humanos de Meta. Aporta conocimiento externo y diferentes puntos de vista y opiniones a nuestros procesos de desarrollo de productos y políticas de contenido y contribuye a brindar transparencia y responsabilidad. Los grupos con quienes interactuamos van desde organizaciones internacionales, grupos de la sociedad civil variados y el entorno académico hasta comunidades marginadas y subrepresentadas, así como también usuarios tales como los [adolescentes](#), [padres](#), [madres y cuidadores](#). Recopilamos las perspectivas de todas las aristas de un tema y de todas las regiones del mundo para garantizar el equilibrio.

Además, seguimos trabajando con nuestros [socios de confianza](#) para identificar tendencias, comprender mejor los desafíos de nuestros canales de escalamiento y explorar de qué modo podemos hacer que esos mecanismos sean más efectivos para los socios de la sociedad civil.

El trabajo proactivo con un amplio espectro de partes interesadas es fundamental para el enfoque de derechos humanos de Meta.

Grupos marginados y defensores de los derechos humanos

Reconocemos la importancia de tener interacciones significativas con las partes interesadas de comunidades marginadas, teniendo en cuenta las variaciones dentro de una región o incluso dentro de un país, y procuramos informarnos respecto de dichas diferencias.

Procuramos escuchar a expertos y activistas de los derechos humanos, académicos y otras personas acerca de una amplia variedad de temas y pedimos su asesoramiento al respecto, y les informamos los desarrollos relevantes de Meta. Sus conocimientos nos dan información para el

desarrollo de nuestras políticas de contenido y la aplicación de políticas a fin de fomentar la expresión y la seguridad.

En todo el mundo, los defensores de los derechos humanos que utilizan plataformas digitales como instrumentos para el activismo se enfrentan con frecuencia a intimidación, opresión y desafíos legales como consecuencia de sus esfuerzos online para el fomento y la protección de los derechos humanos. Esto se nota en particular entre grupos marginados, como las mujeres y la comunidad LGBTQIA+.

Usando nuestro [marco de inclusividad](#) como guía, empleamos mesas redondas, talleres y reuniones individuales para comprender problemas sobre las políticas de contenido relevantes, identificar vacíos normativos y aprovechar oportunidades para la interacción en el desarrollo de políticas. Por ejemplo, trabajamos con [Minority Rights Group](#) para llevar a cabo ocho convocatorias globales en países en riesgo de conflicto armado o violencia social.

También contactamos a más de 250 partes interesadas de diferentes trasfondos, lo que incluyó mujeres, miembros de la comunidad LGBTQIA+,

minorías religiosas, minorías étnicas y grupos indígenas, para recopilar información valiosa sobre los desafíos que enfrentan estas comunidades en materia de políticas. Basándonos en los comentarios obtenidos en más de 19 consultas y talleres con organizaciones LGBTQIA+ y defensores de los derechos humanos en Oriente Medio y el norte de África, colaboramos con la Jordan Open Source Association (JOSA) para lanzar el [kit de herramientas de seguridad digital para defensores de los derechos humanos](#). Esta guía, disponible en [árabe](#) y en [inglés](#), identifica las principales prácticas en seguridad digital y ofrece funciones de seguridad, consejos y medidas de reacción para mejorar la seguridad online de activistas y defensores de los derechos humanos en Oriente Medio y el norte de África. Además, lanzamos una campaña con influencers LGBTQIA+ en África subsahariana con el fin de generar conciencia sobre los recursos de seguridad para activistas LGBTQIA+ y destacar nuestro compromiso con la protección de los usuarios de alto riesgo en la región.

Tras el éxito del piloto centrado en Asia-Pacífico durante dos años, en 2023 expandimos nuestro [Fondo para defensores](#)

[de los derechos humanos](#) que funciona en asociación con [Civil Rights Defenders](#) para apoyar a los defensores de los derechos humanos en todo el mundo y renovamos la financiación con 500.000 USD adicionales que se aplicarán en 2024.

Socios de confianza

AfricanDefenders

A partir de los aprendizajes y los principios del Fondo para la defensa de los derechos humanos en [Asia-Pacífico](#) en asociación con [AfricanDefenders](#), lanzamos el [Fondo para la defensa de los derechos humanos en África](#), diseñado para apoyar a los defensores de los derechos humanos que son objeto de acoso, persecución o actuaciones judiciales debido a sus actividades online en defensa de los derechos humanos. Eso incluye a miembros de grupos marginados que defienden sus derechos, como activistas y periodistas ciudadanos, activistas políticos no violentos, mujeres que defienden los derechos humanos y la comunidad LGBTQIA+. Las microayudas se utilizarán para brindar asistencia en situaciones de emergencia y pagar nuevos dispositivos y tecnologías de seguridad, ofrecer reubicación temporal y brindar asistencia en asuntos jurídicos y de seguridad en situaciones de emergencia.

En enero de 2023, nuestra red de [socios de confianza](#) incluía a más de 400 organizaciones no gubernamentales, agencias humanitarias, organizaciones de defensa de los derechos humanos e investigadores de 113 países de todo el mundo. Además de nuestra participación con partes interesadas continua, extensa y diversificada, estos socios aportan muchísimos conocimientos y experiencia que sirven para dar forma a la moderación de contenido y al desarrollo de políticas efectivas y transparentes. También tienen acceso a un canal de escalamiento para reportar contenido potencialmente dañino y problemas de seguridad de las cuentas, sobre los que actuamos según corresponda para ayudar a mantener la seguridad de los usuarios.

En 2023, los reportes de nuestros socios de confianza nos ayudaron a eliminar 49.600 contenidos que infringían nuestras políticas. Nuestros socios de confianza también nos ayudaron a descubrir y eliminar seis redes engañosas involucradas en [comportamiento no auténtico coordinado](#) en EE. UU., Georgia, Burkina Faso y Togo. Además, tomamos medidas contra tres operaciones independientes de [ciberespionaje en el sur de Asia](#). Estas redes estaban diseñadas para manipular el debate público con un objetivo estratégico, que puede representar una amenaza para los derechos humanos.

En 2023, llevamos a cabo una revisión exhaustiva de nuestros canales para recibir escalamientos relacionados con problemas de seguridad de las cuentas y contenido en situaciones de emergencia de socios de la sociedad civil. Una parte clave de este trabajo fue la organización de una cumbre mundial a la que asistieron expertos en asuntos jurídicos, operaciones y políticas de Meta, junto con 15 socios de la sociedad civil de 13 países de Asia-Pacífico, Oriente Medio, Europa, América Latina y África. A través de esta iniciativa, pudimos evaluar de forma integral los canales de escalamiento para la sociedad civil y formular una visión compartida para su futuro: mejorar la priorización de los reportes y fortalecer los procesos de respuesta ante emergencias y la recopilación de estadísticas.





Caso práctico: mostrar los nuevos daños en Haití

El 7 de julio de 2023 marcó el segundo aniversario del asesinato del presidente de Haití Jovenel Moïse. En el país sigue habiendo mucha tensión, exacerbada por el resurgimiento de la violencia de pandillas. Nuestros socios de confianza reportaron contenido relacionado con varias áreas de infracción, incluidas violencia e incitación, personas y organizaciones peligrosas, bullying y acoso y violencia gráfica. También nos alertaron respecto de la presencia de un nuevo movimiento justiciero, Bwa Kale, en nuestras plataformas. Esta información permitió a nuestros equipos de investigación y operaciones aprovechar las palabras clave e identificar de forma proactiva el contenido que podría infringir las normas y mitigar con eficacia el daño en la vida real.

Caso práctico: aumentar los reportes en Etiopía

Debido a las mayores tensiones sociopolíticas y a la violencia constante en Etiopía, nos reunimos con nuestros socios de confianza de forma mensual para analizar las tendencias de contenido y reforzar las prácticas para reportar contenido de gravedad

alta a través del canal de socios de confianza. También llevamos a cabo una serie de diálogos bimestrales sobre Etiopía con la sociedad civil de Etiopía y la diáspora. Nuestros socios en Etiopía continuaron reportando contenido y proporcionaron evaluaciones de daños en materia de información errónea y reportes de amenazas encubiertas. Al trabajar con estos socios, pudimos identificar mejor algunos problemas clave, incluidas imágenes fuera de contexto, cuentas falsas que difundían información errónea o falsa, cuentas verificadas que utilizaban influencers para difundir información errónea, lenguaje que incita al odio, acoso coordinado y publicación de información personal ("doxing"), términos con connotación política inicialmente acuñados por figuras públicas, y personas y organizaciones peligrosas.

Caso práctico: mitigar los conflictos interreligiosos en Bangladesh

En marzo de 2023, se dio una serie de disturbios interreligiosos entre musulmanes sunitas y ahmadíes en Bangladesh, una comunidad musulmana minoritaria. Nuestros socios de

confianza en Bangladesh reportaron contenido que incluía información errónea, lenguaje que incita al odio y violencia e incitación dirigidos a la comunidad ahmadí. Gracias a las afirmaciones dañinas designadas y revisadas anteriormente, pudimos aplicar las políticas a los reportes de información errónea con rapidez. Los socios de confianza también proporcionaron algunos indicios locales fundamentales que nos permitieron eliminar contenido de acuerdo con políticas exclusivas de la etapa de escalamiento, incluidas amenazas encubiertas, así como también estadísticas sobre términos relacionados con el lenguaje que incita al odio.

Organizaciones internacionales y de varias partes interesadas

A medida que la tecnología avanza y su impacto en los derechos humanos aumenta, seguimos trabajando con una amplia variedad de organismos gubernamentales internacionales, incluidos los siguientes:

- La [Oficina del Alto Comisionado para los Derechos Humanos](#)
- [UNESCO](#)
- La [Oficina para la Prevención del Genocidio y la Responsabilidad de Proteger](#)
- La [Oficina del Enviado del Secretario General para la Tecnología](#)
- [UNICEF](#)
- [ACNUR](#)
- El [Foro para la Gobernanza de Internet](#)
- La [Fundación de las Naciones Unidas](#)
- [OCDE](#)
- El [Foro Económico Mundial](#)
- Diversas organizaciones regionales



Somos miembro de varias iniciativas multisectoriales, como el [Global Internet Forum to Counter Terrorism](#), la [Global Network Initiative](#) y la [Freedom Online Coalition Advisory Network](#). También participamos de forma proactiva en colaboraciones de todo el sector para brindar nuestro aporte y apoyar políticas gubernamentales.

Participamos en la segunda [Cumbre por la Democracia](#) y compartimos públicamente los compromisos en respuesta a la exhortación de la cumbre al sector privado a impulsar la democracia. Nos reconocieron por nuestro coliderazgo en cuanto al desarrollo de principios de la industria para "combatir las amenazas en ascenso por parte de los cibermercenarios" y proporcionamos ideas para la exhortación al sector privado a apoyar [tecnologías anticensura que garanticen la democracia a fin de combatir a los regímenes autoritarios](#). Además, nos reconocieron por nuestro trabajo para posibilitar las conexiones privadas y seguras y contrarrestar el uso indebido de la tecnología digital por parte del Gobierno mediante WhatsApp por proxy y [vigilancia por encargo](#).



Participamos en consultas sobre el desarrollo del [Pacto Digital Mundial](#) con la [Oficina del Enviado del Secretario General para la Tecnología](#) para apoyar la necesidad de contar con un enfoque internacional armonioso respecto de la gobernanza de internet y proteger los derechos humanos online, incluidos los derechos de privacidad y libertad de expresión. También participamos en consultas sobre el [Código para la integridad de la información de la ONU](#) y proporcionamos una solicitud detallada.

Seguimos manteniendo reuniones periódicas con funcionarios de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, expertos independientes y relatores especiales para debatir asuntos mundiales y específicos de cada país, y participar de manera activa en el [Proyecto B-Tech](#). Hicimos un pacto de compromiso con los derechos humanos en ocasión del 75.º aniversario de la Declaración Universal de los Derechos Humanos. Esto incluyó la renovación de compromisos con la divulgación y la debida diligencia en materia de derechos humanos y el trabajo con la sociedad civil y las comunidades afectadas, entre otras cuestiones.

Meta formó parte de la [iniciativa colaborativa Por un internet confiable](#) de la UNESCO, y participó de la cumbre de febrero en París y del proceso multisectorial para desarrollar pautas basadas en los derechos humanos sobre las regulaciones de plataformas digitales.

En colaboración con el [Foro Económico Mundial](#), participamos en un proyecto amplio para generar consenso sobre políticas relacionadas con la tecnología emergente. Ayudamos a desarrollar y avalamos los [Principios globales sobre seguridad digital: interpretar los derechos humanos internacionales en la era digital](#), que establecen un marco diseñado para ayudar a los Gobiernos y proveedores de servicios online a impulsar la seguridad digital a través de un enfoque multisectorial. También somos parte de la [AI Governance Alliance](#) y de la iniciativa para [definir y](#)

"No podemos hacerlo solos. Necesitamos tener interacciones estratégicas y asociaciones estratégicas para que este proyecto tenga un impacto".

Lene Wendland, jefa de la Unidad de Empresas y Derechos Humanos, Derechos Humanos de la ONU

[construir el metaverso](#), que busca guiar el desarrollo de un metaverso responsable y ético. El año pasado, nos incorporamos a la [AI Governance Alliance](#), que busca establecer normas mundiales para la gobernanza de la IA, de manera de contribuir de forma positiva al avance tecnológico para el bienestar social, guiados por compromisos con la ética y la diversidad.

Comenzamos a trabajar con la [Comisión de Estupefacientes de la ONU](#) para establecer The Prevent Alliance, una iniciativa público-privada entre Meta, Snap, el Gobierno de EE. UU. y la Oficina de las Naciones Unidas contra la Droga y el Delito dirigida a evitar el uso indebido de plataformas digitales para realizar actividades dañinas e ilícitas relacionadas con el uso no medicinal de drogas sintéticas.

Nos [comprometimos](#) con el [Foro Mundial sobre los Refugiados](#) para apoyar los derechos de los refugiados, los solicitantes de asilo y las personas desplazadas internas.

Data for Good at Meta

Nuestro programa Data for Good sigue trabajando con agencias humanitarias, organizaciones sin fines de lucro, investigadores y Gobiernos de todo el mundo para apoyar a las poblaciones marginadas. Por ejemplo, este programa sigue apoyando la labor de las organizaciones sin fines de lucro como las iniciativas de IMPACT que ayudan a reubicar a quienes fueron desplazados por la guerra en Ucrania. También apoya a agencias internacionales, incluida la Organización Internacional para las Migraciones en lo relativo a [métricas mundiales](#) mejoradas para apoyar a las poblaciones migrantes a nivel mundial y aquellos desplazados

por desastres naturales. Además, Data for Good at Meta contribuye muchísimo al trabajo de instituciones de desarrollo, como el Banco Mundial, a través de colaboraciones a gran escala como la Development Data Partnership, que permite que grandes instituciones multilaterales orienten mejor aquellos programas que ayuden a las comunidades que lo necesitan. Por ejemplo, las herramientas basadas en IA creadas por Data for Good, como el índice de riqueza relativa, han ayudado al Banco Mundial a comprender mejor de qué manera la contaminación del aire afecta a las poblaciones más pobres en países de ingresos bajos y medios.





Transparencia y reparación



Estamos comprometidos con la transparencia y la reparación como principios rectores de la gestión de riesgos para los derechos humanos. Ambos son elementos esenciales consagrados en los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#). Nuestros [informes de transparencia](#), que se publican de forma periódica, y nuestro informe anual sobre derechos humanos muestran cómo aplicamos las [Normas comunitarias de Facebook](#) y las [Normas comunitarias de Instagram](#), respondemos a las solicitudes de los Gobiernos y protegemos la propiedad intelectual. También trabajamos mucho para seguir mejorando la accesibilidad de nuestras Normas comunitarias, que estaban disponibles en 90 idiomas a finales de 2023. Creemos que el Consejo asesor de contenido continúa proporcionando acceso a un recurso único en el sector de medios sociales.



Responder a solicitudes gubernamentales

Riesgos significativos relevantes:

- Libertad de opinión y expresión
- Privacidad
- Derecho a un juicio justo
- Derecho a un recurso efectivo

No tomamos ninguna medida sobre el contenido ni divulgamos datos de ningún usuario solo porque una entidad gubernamental nos lo solicite. Como miembros de la [Global Network Initiative](#) (GNI), nos comprometemos con los principios de libertad de expresión y privacidad de la GNI (principios de la GNI), que se basan en normas de derechos humanos reconocidos a nivel internacional. Esto es importante porque dichos principios nos brindan orientación sobre la manera de responder a las solicitudes gubernamentales para minimizar los impactos negativos sobre la libertad de expresión y la privacidad de los usuarios.

Con el correr de los años, las solicitudes gubernamentales fueron cambiando. Los agentes gubernamentales nos solicitan que adoptemos una amplia variedad de medidas más allá de restringir el acceso a un determinado contenido en un país. Desde 2020, la cantidad de solicitudes gubernamentales aumentó y también aumentó la cantidad de contenido que Meta restringe en virtud de la legislación local.

Diez años con la Global Network Initiative

"Los principios de la GNI son nuestra guía para administrar las solicitudes de eliminación de los Gobiernos. Las solicitudes para restringir contenido basadas en la legislación local suelen plantear dilemas trascendentales y muy reales con respecto a los derechos humanos. Cuando nos enfrentamos a conflictos entre la legislación local y nuestros compromisos con los derechos humanos, recurrimos a la guía de la GNI, las normas de derechos humanos reconocidas internacionalmente y nuestra Política corporativa derechos humanos".

Siobhán Cummiskey, directora de Política de Contenido, Meta

La [GNI](#) es una iniciativa multisectorial, cuyos miembros se comprometen a defender los derechos de libertad de expresión y privacidad según se definen en un tratado clave sobre los derechos humanos: el Pacto Internacional de Derechos Civiles y Políticos (ICCPR). La GNI establece un criterio internacional para la toma de decisiones responsables por parte de las empresas para promover y fomentar los derechos de privacidad y libertad de expresión en todo el ecosistema tecnológico. En el marco de los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas, los principios de la GNI actúan como un enlace entre las obligaciones de los Gobiernos de

proteger los derechos humanos y la responsabilidad de las empresas de respetar esos derechos.

El 2023 marcó el décimo aniversario de la incorporación de Meta a la GNI y de nuestro compromiso con los derechos humanos. Aplicamos los principios de la GNI (y tenemos la responsabilidad de cumplirlos) en lo relativo a la manera en que las empresas de tecnología deben respetar los derechos de privacidad y libertad de expresión de sus usuarios a la hora de abordar solicitudes y restricciones gubernamentales. La manera en que realizamos cualquier implementación se evalúa de forma independiente y periódica.



Las solicitudes gubernamentales abarcan tanto restricciones de contenido como acceso a datos de usuarios. Estas solicitudes afectan los derechos de libertad de expresión y privacidad, respectivamente. Cuando recibimos un reporte en el que se nos pide restringir contenido o una solicitud de acceso a datos de usuarios por parte de una entidad gubernamental o un tribunal, lo analizamos siguiendo un proceso global uniforme para ambas situaciones (consulta [aquí](#) y [aquí](#)).

- Los funcionarios del Gobierno suelen solicitar datos de personas que utilizan nuestras plataformas en el marco de investigaciones oficiales. Seguimos aplicando los [principios de la GNI](#) a nuestras respuestas para todo tipo de solicitudes gubernamentales. La gran mayoría de estas solicitudes se relacionan con causas penales. En muchos de estos casos, el objetivo de estas solicitudes gubernamentales es obtener información básica de los suscriptores, como el nombre, la fecha de registro y la duración del servicio. En otras solicitudes, el objetivo es obtener los registros de la dirección IP o el contenido de las cuentas. Aplicamos [normas](#) estrictas para gestionar todas las solicitudes gubernamentales de datos. Informamos el número de solicitudes recibidas, el número de usuarios o cuentas solicitadas y el porcentaje de solicitudes en las que proporcionamos datos. [Informamos](#) públicamente estos datos junto con nuestras políticas para responder a las solicitudes desde 2016.
- En ocasiones, las autoridades gubernamentales solicitan la eliminación de contenido por presunta infracción de la legislación local. También podemos recibir una orden judicial donde se nos exija la eliminación de contenido. Si determinamos que el contenido no infringe nuestras políticas, podemos restringir el acceso al contenido en el país donde se alega que es ilícito después de realizar una evaluación legal y de derechos humanos congruente con nuestra [Política corporativa de derechos humanos](#) y nuestros compromisos asumidos

como miembro de la GNI. También informamos al usuario afectado, siempre que la ley no nos lo prohíba. Consideramos si debe implementarse alguna medida de mitigación, como realizar una serie de acciones para limitar el impacto de la solicitud, lo que incluye cumplir bajo protesta (que en algunos casos ha dado lugar a la revocación de las solicitudes en cuestión), o la presentación de una apelación legal a la solicitud de eliminación por parte de Meta, transparencia en tiempo real a través de la publicación de un caso práctico en el [Centro de transparencia](#) o el envío de una copia de la solicitud a la [base de datos de Lumen](#). Si consideramos que las solicitudes gubernamentales u órdenes judiciales no son válidas desde un punto de vista jurídico, o que son demasiado generales o no son congruentes con las normas internacionales de derechos humanos, podemos solicitar una aclaración, apelar o no tomar ninguna medida.

- En ocasiones recibimos solicitudes gubernamentales u órdenes judiciales donde se nos pide que adoptemos medidas especiales en virtud de requisitos establecidos en la legislación local, lo que incluye aplicar un aviso de corrección en el contenido (p. ej., que indique que las autoridades locales consideran que el contenido constituye información errónea), limitar el acceso de un usuario a funciones específicas, restablecer contenido eliminado anteriormente por infringir nuestras políticas, la obligación de restringir automáticamente el contenido a escala y en el país, y restringir el contenido en todo el mundo, aunque solo infrinja la legislación local de un único país. Si bien respetamos la legislación de los países donde operamos, cuando nos encontramos con conflictos entre las obligaciones legales nacionales y nuestros compromisos con los derechos humanos y la transparencia, procuramos respetar los principios de los derechos humanos reconocidos internacionalmente en la máxima medida posible. Podemos solicitar una aclaración, rechazar la solicitud o no tomar ninguna medida.



- En algunos casos es posible que tengamos prohibido por ley publicar solicitudes de eliminación o determinada información incluida en ellas. Cuando esto suceda, procuraremos publicar tanta información como sea posible sobre la solicitud y su existencia sin infringir nuestras obligaciones legales.

Cómo leer los informes de restricción de contenido

Recibimos reportes de Gobiernos, autoridades reguladoras y tribunales, así como de entidades no gubernamentales y miembros del público, en relación con contenido que presuntamente infringe la legislación local. El [informe de restricción de contenido](#), que se publica semestralmente desde 2013, detalla los casos en que restringimos el acceso a contenido en Facebook e Instagram en virtud de la legislación local.

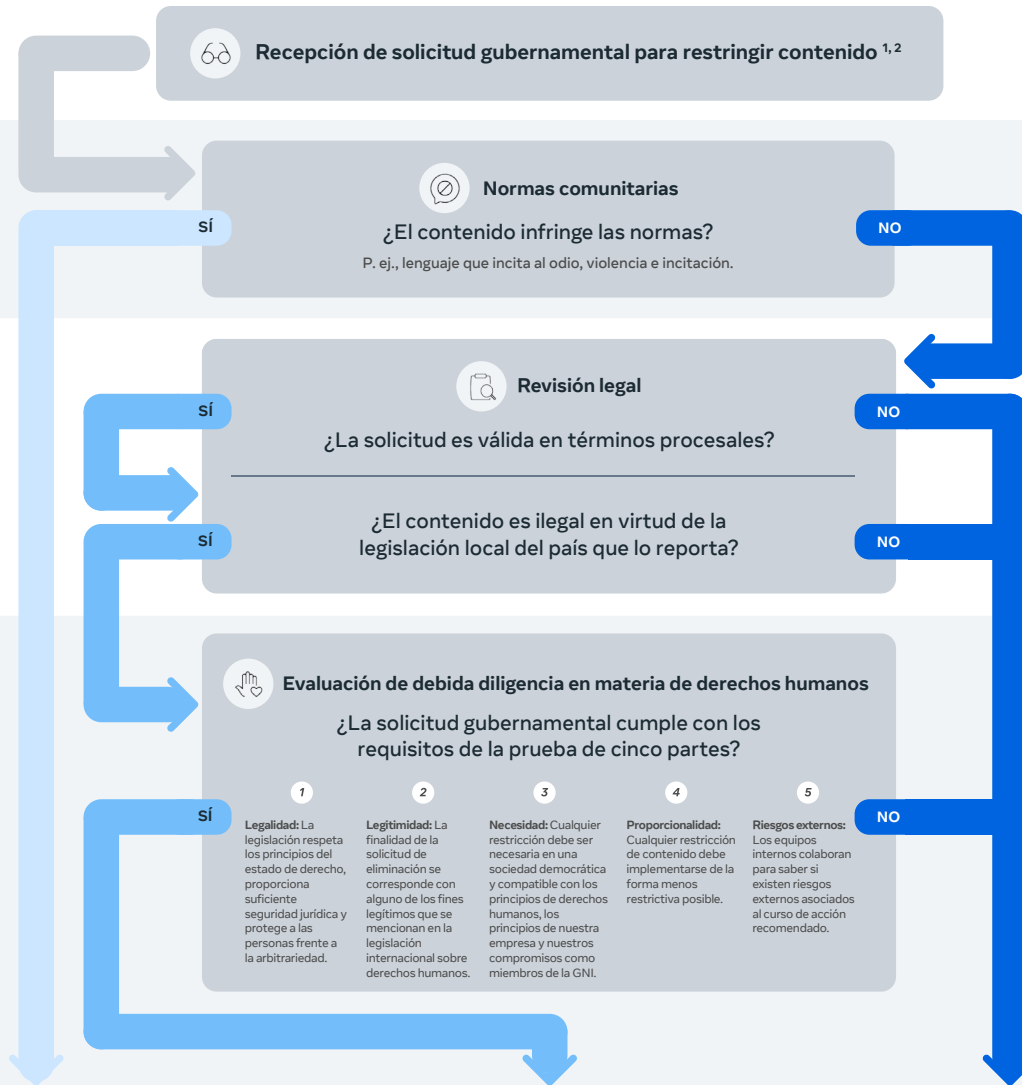
- La sección de [datos específicos por país](#) proporciona información detallada sobre la cantidad, la naturaleza y el tipo de contenido que restringimos cada seis meses en virtud de la legislación local. Desde 2023, nuestros informes también incluyen información sobre determinados países donde la legislación local nos obliga a restringir automáticamente el contenido, a escala y en el país, o a aplicar un aviso de corrección del contenido en el país.
- La sección de [restricciones internacionales](#) detalla los casos infrecuentes en que tenemos la obligación legal de restringir la disponibilidad de contenido en todo el mundo debido a exigencias legales extraterritoriales, aunque el contenido en cuestión solo infrinja la legislación local de un país específico.

- La sección de [casos prácticos](#) proporciona información detallada sobre algunas de las solicitudes de eliminación de contenido que recibimos por parte de Gobiernos y tribunales. Desde 2023, nos centramos en publicar solo los casos prácticos con un alto valor de interés público.
- La sección de [evaluación de informes de contenido que infringe la legislación local](#) contiene información sobre nuestro proceso de revisión de solicitudes de eliminación gubernamentales y las medidas de mitigación que adoptamos para limitar el impacto de las restricciones en consonancia con las normas internacionales de derechos humanos.

Publicamos los datos sobre las restricciones en virtud de la legislación local para el segundo semestre de 2023 en mayo de 2024 (disponible [aquí](#)).



¿Cuál es el ciclo de vida de una solicitud gubernamental?



Resultado

Podemos implementar medidas de mitigación, como presentar apelaciones o compartir información sobre la solicitud gubernamental para restringir el contenido externamente a fin de limitar su impacto.



Medidas conforme a nuestras políticas

En el caso de que identifiquemos contenido que infringe las Normas comunitarias de Facebook o las Normas comunitarias de Instagram, tomaremos las medidas que correspondan (p. ej., eliminaremos el contenido o lo restringiremos por edad, etc.).



Restringir el acceso en jurisdicciones donde es ilegal

Informamos a las personas cuando restringimos contenido que publicaron basándonos en un reporte donde se alegue que ese contenido infringe la legislación local. Además, informamos a las personas cuando intentan ver contenido que se restringió en su país debido a una solicitud gubernamental para restringir contenido. También notificamos a la autoridad gubernamental relevante cuando tomamos medidas en respuesta a su solicitud.



No tomar ninguna medida O solicitar información adicional

Podemos contactarnos con la autoridad correspondiente para solicitar información adicional sobre la solicitud.

¹ Esta ilustración refleja el proceso para revisar y abordar solicitudes gubernamentales donde se nos pide que restrinjamos contenido en Facebook e Instagram según se describe en el Centro de transparencia. Para acceder a una descripción más detallada, visita el Centro de transparencia (a través de este enlace). Puede haber algunas diferencias para los anuncios y otros productos de Meta.

² Es posible que tengamos que apartarnos del ciclo de vida descrito arriba durante las situaciones de emergencia. Aunque no es habitual, en ocasiones la legislación de un país puede obligarnos a restringir automáticamente el acceso a determinado contenido, a escala en países específicos, en virtud de los requisitos de la legislación local. En dichos casos, seguiremos basando nuestro enfoque en los compromisos asumidos en virtud de la Global Network Initiative y nuestra Política corporativa de derechos humanos.

Actualización de nuestra participación en Lumen

En marzo de 2022, nos comprometimos a [participar en Lumen](#), un proyecto de investigación independiente a cargo del Berkman Klein Center for Internet and Society de la Universidad de Harvard. Este proyecto permite a los investigadores analizar las solicitudes de eliminación de contenido efectuadas por Gobiernos y agentes privados en relación con contenido online. En noviembre de 2023, enviamos nuestra primera serie de solicitudes de eliminación de Austria, Vietnam, Singapur, India y México a [Lumen](#), a las que se puede acceder a través de su base de datos. Esto facilitará las iniciativas de la comunidad internacional a la hora de realizar análisis y defender los derechos digitales de los usuarios de internet y preparar informes al respecto. Los principios en los que basamos nuestro enfoque a la hora de compartir información públicamente acerca de solicitudes de eliminación gubernamentales pueden consultarse en nuestros [informes de restricciones de contenido](#).

Aviso mejorado para los usuarios

Para mejorar aún más la transparencia en consonancia con el derecho a un recurso efectivo y el derecho a un juicio justo, en 2023 mejoramos nuestro aviso para los usuarios en Facebook e Instagram respecto del contenido restringido por Meta en virtud de solicitudes legales por parte de Gobiernos y tribunales. En la mayoría de los casos, mediante nuestras notificaciones informamos a los usuarios qué autoridad estatal envió la solicitud que generó la restricción y el país en que se restringió el contenido. No se aplica al contenido que eliminamos por infringir nuestras políticas.





Riesgos significativos relevantes:

- Derecho a un recurso efectivo
- Derecho a un juicio justo
- Libertad de opinión y expresión
- Vida, libertad y seguridad de la persona
- Derecho a la participación pública, al voto y a presentarse a una candidatura
- Igualdad y no discriminación
- Interés superior del niño
- Privacidad

Consejo asesor de contenido

El Consejo asesor de contenido es una iniciativa líder en el sector creada para ayudar a Meta a abordar algunas de las cuestiones más complejas en torno a la libertad de expresión y a la seguridad online, qué se debe eliminar, que se debe conservar y por qué. Es un órgano independiente que puede revisar casos tanto de Meta directamente como de personas que usan Facebook, Instagram o Threads que no están de acuerdo con las decisiones que tomamos con respecto a la moderación de contenido. Además de pronunciar decisiones vinculantes respecto de si un contenido debe conservarse o eliminarse, el Consejo publica recomendaciones para ayudar a mejorar la moderación de contenido, así como opiniones de asesoramiento normativo cuando lo solicitamos. El Consejo asesor de contenido aplica las normas internacionales de derechos humanos para proporcionar a Meta información valiosa y recomendaciones de expertos a fin de ayudar a garantizar que las decisiones sobre nuestras políticas y productos se tomen en beneficio de los usuarios.

El año 2023 fue un año histórico para el Consejo asesor de contenido, que superó su objetivo de 50 decisiones por año, más del triple de las decisiones tomadas en 2022.

El Consejo asesor de contenido realizó cambios importantes a su estatuto, con el objeto de contemplar [revisiones urgentes](#) en casos excepcionales, incluido cuando el contenido podría generar consecuencias graves en el mundo real. Este proceso permite que se emitan decisiones sobre contenido aceleradas en un plazo máximo de 30 días. El Consejo utilizó el proceso de revisión acelerada por primera vez para dos casos relacionados con el [conflicto entre Israel y Hamás](#).

Las decisiones del Consejo abarcaron diferentes problemáticas en un gran número de países, incluido el eslogan de protesta de Irán "Muerte a Khamenei", el caso del primer ministro de Camboya Hun Sen, cuestiones de género y desnudos, prisioneros de guerra armenios, las elecciones en Brasil y protestas de mujeres en Cuba, entre otros casos. Además de sus decisiones, el Consejo le hizo 60 recomendaciones a Meta en 2023. También publicó una opinión de asesoramiento normativo respecto de la eliminación de información errónea sobre COVID-19. Para obtener más información, consulta los informes de transparencia del Consejo asesor de contenido disponibles [aquí](#), [aquí](#) y [aquí](#).

Concesión de interés periodístico

En respuesta a las recomendaciones del Consejo asesor de contenido, proporcionamos mayor transparencia respecto de la manera y las situaciones en que aplicamos la concesión de interés periodístico al contenido que podría infringir las Normas comunitarias de Facebook o las Normas comunitarias de Instagram, pero que mantenemos visible por una cuestión de interés público.

- Entre el 1 de junio de 2022 y el 1 de junio de 2023, documentamos 69 concesiones de interés periodístico.
- De las concesiones documentadas, nueve (~13%) de ellas se otorgaron a publicaciones de políticos.
- De las 69 concesiones, documentamos un total de 17 concesiones amplias, es decir que cada una se aplica a más de un contenido.



Programa de verificación cruzada

El 2023 también marcó un punto de inflexión importante para nuestro programa de verificación cruzada, ya que comenzó con nuestra respuesta inicial a las 33 recomendaciones del Consejo asesor de contenido relativas a la sofisticación técnica, el rigor operativo, la igualdad, la gobernanza y la transparencia del programa. Implementamos acciones para aumentar la participación de nuestros socios de confianza y de la sociedad civil en el desarrollo de la gobernanza de verificación cruzada mediante sesiones de comentarios con socios de nueve regiones diferentes.



En 2023, de las 122 recomendaciones en las que trabajamos, completamos un total de 61 y seguimos trabajando en muchas más (para obtener información detallada, consulta las [actualizaciones semestrales del Consejo asesor de contenido](#)). Las recomendaciones que aplicamos abarcaron nuestras operaciones, políticas y productos, lo que contribuyó a conseguir mejoras amplias y significativas en toda la empresa y en nuestra comunidad internacional.

Recomendaciones del Consejo asesor de contenido en 2023

Recomendaciones del Consejo asesor de contenido

66

(91 en 2022)

Evaluación o implementación de Meta en curso*

69

(75 en 2022)

Recomendaciones implementadas en su totalidad*

61

(14 en 2022)

* Algunas evaluaciones o implementaciones en curso o recomendaciones implementadas en su totalidad incluyen recomendaciones de años anteriores (consulta el [informe sobre derechos humanos de 2022](#) para obtener más información).



Consideraciones para el futuro

En 2023, fuimos testigos de muchos desarrollos a un ritmo acelerado, como se describe en este informe, y de muchas oportunidades de incluir los derechos humanos en la toma de decisiones. Procuramos hacerlo. Sin embargo, los contextos sociopolíticos en general son inestables, y muchas fuentes de datos muestran una erosión de las normas democráticas y de libertad de expresión. Las tecnologías de inteligencia artificial (IA) existen hace décadas, pero la conciencia pública sobre las oportunidades y los riesgos que estas pueden generar se disparó en 2023.

El marco regulatorio en rápida evolución es a la vez prometedor y complejo para los derechos humanos: puede impulsar la innovación respecto de la manera, y en qué medida, las empresas gestionan los riesgos relacionados con los derechos humanos de acuerdo con los principios clave de derechos humanos, incluidos los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas. Sin embargo, las exigencias en materia de cumplimiento también podrían apagar el deseo de innovar y diluir el margen para hacerlo. Es cada vez más evidente que una gestión continua de los riesgos para los derechos humanos a escala implica el uso constante de herramientas y datos que se ajusten a los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas

Meta procura innovar en sus metodologías de gestión de riesgos para los derechos humanos, lo que incluye nuestras interacciones con las partes interesadas. El aprendizaje basado en la experiencia, los ejercicios de

"Estamos repletos de oportunidades para que las empresas y los Gobiernos tengan en cuenta los derechos humanos en sus decisiones cotidianas".

John Ruggie, autor de los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas

simulación y el estudio de casos prácticos parecen ser prácticas recomendadas prometedoras, pero aún falta para que se conviertan en algo generalizado y procuramos afianzar nuestro trabajo en estas áreas. Como se analizó en este informe, también reconocemos la importancia de implementar iniciativas e innovaciones coordinadas a la hora de abordar los riesgos potenciales relacionados con los derechos humanos vinculados con la IA generativa.

Sabemos que aún queda mucho trabajo por hacer. Queremos construir a partir de nuestro trabajo y las lecciones aprendidas para contribuir a la evolución del conocimiento y las buenas prácticas para el sector de la tecnología y las comunidades de derechos humanos y empresas.

Estamos en un momento tanto complejo como prometedor para los derechos humanos en el sector tecnológico. Todos nosotros, incluidas las empresas, la sociedad civil, los Gobiernos, el sistema de la ONU, los inversores y las autoridades reguladoras,



deberemos trabajar para desarrollar nuevas habilidades y nuevo conocimiento, junto con nuevas maneras de colaborar, para diseñar, compartir y aprender la mejor manera de gestionar los riesgos para los derechos humanos a escala.

Se acercan futuras olas de inmensa creatividad: necesitaremos de muchas personas valientes, experimentos e intervenciones de emprendedores para aprovecharlas. Afortunadamente, al movimiento de derechos humanos nunca le ha faltado la valentía para innovar. ¡Manos a la obra!



Anexo

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

Para avanzar en nuestra labor para respetar los derechos humanos en nuestros programas, servicios y políticas, debemos contar con estructuras de administración y gobernanza claras. Nuestros expertos en derechos humanos guían 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.

Las tareas de los expertos en derechos humanos 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 proporciona orientación 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 encarga 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 para los derechos humanos, así como las acciones que la gerencia llevó a cabo para controlar o mitigar dicha exposición. El comité recibe información de manera periódica sobre la labor actual y en curso que realizan los expertos en derechos humanos.

Capacitación en materia de derechos humanos 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 es la capacitación que lanzamos en 2022 y continuó a lo largo de 2023. Destaca el impacto potencial y real de nuestros servicios, políticas y decisiones comerciales 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 se centra en los principios de justicia, igualdad y no discriminación.

Nuestros objetivos de capacitación sobre derechos humanos también están respaldados por nuestra capacitación anual obligatoria sobre privacidad. Esta se centra en desarrollar nuestra capacidad colectiva para proteger a las personas, incluidas, en especial, las categorías de personas marginadas, frente a daños que surjan del tratamiento de datos de personas. Esta capacitación aborda la protección del derecho de privacidad de las personas y la protección de datos.



Enlaces a los informes mencionados

- [Informe de sostenibilidad de 2024](#)
- [Informe de prácticas comerciales responsables de 2024](#)
- [Informe sobre derechos humanos de 2022, Informe sobre derechos humanos de 2021](#)
- [Informe sobre el combate de la esclavitud y la trata de personas de 2023](#)
- [Informe sobre el combate de la esclavitud y la trata de personas de 2024](#)
- [Informe sobre minerales en zonas de conflicto de 2023](#)
- [Actualización del progreso en materia de derechos civiles de 2023](#)
- [Informes de transparencia de Meta](#)
- [Informes regulatorios y otros informes de transparencia](#)
- Evaluaciones sobre el impacto en los derechos humanos: [Cifrado de extremo a extremo](#), [Filipinas](#), [Myanmar](#), [Indonesia](#), [Camboya](#), [India](#), [Sri Lanka](#) e [Israel y Palestina](#)

