

Informe sobre derechos humanos



Índice

Acerca de este informe	03
Resumen ejecutivo	06
Gestión de riesgos para los derechos humanos	09
1. Libertad de opinión y expresión	10
2. Privacidad	13
3. Igualdad y no discriminación	14
4. Vida, libertad y seguridad de la persona	15
5. Interés superior del niño	16
6. Derecho a la participación pública, al voto y a presentarse a una candidatura	16
7. Libertad de reunión y asociación	17
8. Derecho a la salud	17
Temas destacados	18
Innovar respetando los derechos humanos	18
Mejorar la accesibilidad con la tecnología de dispositivos portátiles	19
Seguir innovando de manera responsable en el campo de la IA	19
Fraudes y estafas	22
Desarticular redes criminales de estafas en el Sudeste Asiático	22
Cómo nos preparamos para afrontar una crisis y respondemos ante estas situaciones	24
República Democrática del Congo	25



Participación de partes interesadas	26
Socios de confianza	27
Caso de éxito: Información valiosa para prevenir el reclutamiento de jóvenes en Latinoamérica	28
Organizaciones internacionales	29
Transparencia y reparación	30
Anexo	33
Cómo se gestionan y regulan los derechos humanos en Meta	34
Capacitación en materia de derechos humanos para empleados de Meta	34
Enlaces a los informes mencionados	34





Acerca de este informe

Este informe anual sobre derechos humanos abarca información valiosa y las medidas que se tomaron desde el 1 de enero hasta el 31 de diciembre de 2025. Incluimos información sobre los servicios y productos de Meta, como Facebook, Messenger, Instagram, WhatsApp, Threads y Reality Labs.

Se basa en el trabajo realizado por Meta en pro de los derechos humanos 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](#). En el informe, se explica cómo aplicamos estos principios en la empresa en 2025 y se indica dónde encontrar información más detallada al respecto.



El contenido de este informe se basa en nuestra [Evaluación integral de riesgos significativos para los derechos humanos](#), que se realizó en 2022. El propósito de la evaluación era identificar y priorizar los impactos¹ potencialmente adversos para los derechos humanos más significativos que afectan a las personas que usan nuestros productos y a otras en quienes repercuten las medidas que adoptamos. En este informe, se detallan estos posibles riesgos significativos y se incluyen ejemplos de las medidas que tomamos y las iniciativas de mitigación que emprendimos en 2025.

En 2025, los derechos humanos seguían siendo un tema de fundamental importancia tanto para nuestra empresa como para nuestras partes interesadas. Abreviamos y simplificamos el informe sin dejar de esforzarnos por proporcionar una muestra representativa de nuestra labor en diversos equipos y en colaboración con partes interesadas de todo el mundo.

Políticas y progreso

Además de este informe acerca de derechos humanos, Meta elabora cada año informes sobre políticas y el progreso alcanzado mediante los siguientes mecanismos:



Informe anual



Declaración de representación



Informe de prácticas comerciales responsables



Centro de transparencia



Informe de sustentabilidad



Informe sobre el cambio climático del CDP



Pacto Mundial de las Naciones Unidas

Este informe complementa el [Informe de prácticas comerciales responsables de Meta](#). Informamos por separado sobre 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. Asimismo, presentamos los informes nacionales y de la Unión Europea que son obligatorios, y que están disponibles en nuestro [Centro de transparencia](#). En el [Anexo](#) de este informe, se incluyen enlaces a otras divulgaciones de Meta.

[Ir al Anexo](#)

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 a cómo repercute en una persona cuando una medida reduce o anula su capacidad de ejercer sus derechos humanos.



Nuestra [Política corporativa de derechos humanos](#) se aplica a toda la empresa. Cada servicio y entidad de Meta cuenta con políticas y procedimientos propios que pueden repercutir de manera diferente en los derechos humanos. 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 o circunstancias.²



² 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 genera, nuevas obligaciones (legales o de otro tipo) respecto de la aplicación de una política o un procedimiento a otros servicios o entidades.



Resumen ejecutivo



Este es el quinto informe anual sobre derechos humanos de Meta. En él, se brinda información valiosa sobre la labor que realizó Meta en 2025 para gestionar los riesgos para los derechos humanos y satisfacer los compromisos que asumimos con los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#).



Evaluación de riesgos significativos

Nuestras prioridades en 2025 reflejaron los riesgos significativos que se identificaron en nuestra [Evaluación integral de riesgos significativos para los derechos humanos](#) de 2022: 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.

[→ Leer más](#)

Impulsar la innovación

Meta buscó ampliar los límites de la innovación en inteligencia artificial (IA) al mismo tiempo que integraba consideraciones sobre los derechos humanos en nuestro proceso de desarrollo. Presentamos nuevos modelos, marcos de seguridad y productos, como [los lentes Meta Ray-Ban Display y Meta Neural Band](#). Esta pulsera usa señales musculares para controlar los lentes con IA sin usar las manos. Desarrollamos estos productos a través de nuestro proceso de desarrollo responsable, teniendo en cuenta la [accesibilidad](#) y respetando el derecho a la igualdad y a la no discriminación.

[→ Leer más](#)

Enfrentar el fraude y las estafas

También ampliamos nuestras iniciativas para combatir el fraude y las estafas. Basamos [estas iniciativas](#) en la aplicación de las normas de la plataforma y en colaboraciones interregionales con miembros del sector, las autoridades y las organizaciones locales de la sociedad civil. Muchos centros de estafas se valen de la trata de personas y el trabajo forzado para llevar a cabo sus actividades. Por consiguiente, nuestras iniciativas destinadas a mitigar el problema ayudaron a garantizar el derecho a la vida, la libertad y la seguridad, así como la prohibición de la esclavitud, la servidumbre y el trabajo forzado.

[→ Leer más](#)

Seguridad infantil y juvenil

De igual manera, continuamos con nuestro compromiso con la [seguridad infantil y juvenil](#). Entre otras iniciativas, anunciamos que Instagram [rediseñaría](#) la experiencia de contenido en las cuentas de adolescentes, inspirándonos en los criterios de clasificación de películas para mayores de 13 años y en los comentarios de los padres.* También desarrollamos esta experiencia actualizada siguiendo las recomendaciones de los especialistas y el principio de las capacidades en desarrollo del niño, tal como se describe en la [Convención sobre los Derechos del Niño de la ONU](#).

[→ Leer más](#)

Respuesta a crisis

En nuestro Informe sobre derechos humanos de 2024, aportamos información detallada sobre cómo Meta integra los principios de derechos humanos en [la forma en que nos preparamos y respondemos ante las crisis](#), siguiendo nuestro [Protocolo de la política de crisis](#). Esta labor continuó en 2025. Aplicamos el Protocolo de la política de crisis para responder a una variedad de situaciones. Por ejemplo, las que se dieron en la India (Cachemira) e Indonesia, y las "protestas de la generación Z" lideradas por jóvenes en Madagascar, Serbia y otros países. En este informe, presentamos un caso de éxito sobre nuestra respuesta ante la crisis en la República Democrática del Congo.

[→ Leer más](#)



Participación de partes interesadas

Nuestra [Política corporativa de derechos humanos](#) respalda nuestra colaboración proactiva con partes interesadas. En 2025, colaboramos con diversos grupos de derechos humanos, comunidades vulnerables, miembros de la sociedad civil, académicos, laboratorios de ideas y autoridades reguladoras. Entre los principales temas, se incluyeron el enfoque de Meta respecto a la IA y la moderación de contenido, el impacto de nuestros [cambios en políticas y productos de enero](#), y nuestras iniciativas para combatir el fraude y las estafas.

Seguimos trabajando con [socios de confianza](#) de todo el mundo para identificar tendencias y comprender mejor el impacto del contenido y el comportamiento online en las comunidades locales. Nuestros socios proporcionaron información y detectaron contenido que podría infringir las normas en más de 20 países y regiones.

→ Leer más

Consejo asesor de contenido

También brindamos información actualizada sobre el trabajo del [Consejo asesor de contenido](#). El Consejo asesor de contenido es un organismo independiente que emite resoluciones vinculantes y recomendaciones no vinculantes. Durante el período que abarca este informe, el Consejo asesor de contenido analizó casos relacionados con nuestras iniciativas para respetar los derechos humanos, como la libertad de expresión, los derechos e intereses de los jóvenes, y el derecho a la igualdad y a la no discriminación, entre otros temas.

A modo de ejemplo, en respuesta a una recomendación, Meta [aumentó](#) el nivel de detalle de las notificaciones existentes. Gracias a esto, los usuarios pueden recibir información más detallada cuando se restringe su contenido, en respuesta a un informe oficial del Gobierno que indica que dicho contenido infringe las leyes locales.

→ Leer más

Administración de solicitudes gubernamentales

A lo largo de 2025, continuamos tomando como guía nuestro compromiso con [Global Network Initiative](#) respecto de la libertad de expresión y la privacidad, incluso al responder a solicitudes gubernamentales para restringir contenido. Estas solicitudes del Gobierno vienen aumentando. Además de nuestros informes habituales sobre integridad, publicamos [casos de éxito](#) relacionados con las solicitudes del Gobierno para restringir la libertad de expresión política en Alemania, Austria, Colombia, Costa de Marfil, India, Irak, Italia, Kazajistán, Malasia, Nueva Zelanda, Polonia, Rumania, Singapur, Sudáfrica, Tanzania, Turquía y Vietnam. En 2025, Meta también se sometió a su cuarta evaluación de GNI (que abarcó el período de abril de 2022 a diciembre de 2024). El Consejo de GNI determinó de manera unánime que Meta había "hecho esfuerzos de buena fe para implementar los Principios de GNI, con una mejora a lo largo del tiempo".

Ver casos de éxito

Gestión de riesgos para los derechos humanos

Los [Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas](#) dejan claro que las empresas deben identificar el impacto negativo que tienen sobre los derechos humanos a fin de evitarlo o mitigarlo de manera efectiva.

Dados el alcance de las operaciones de Meta y el rango de derechos que podría implicar, anticiparse a los [riesgos significativos](#) y abordarlos es una tarea importante, pero compleja. Gestionamos dos tipos de riesgos inherentes: aquellos que proceden de nuestras propias actividades y los que surgen a raíz de las actividades de terceros, incluidas las personas que usan nuestras plataformas.



En la tabla a continuación, se detallan los riesgos significativos inherentes para los derechos humanos³ que detectamos, conforme se definen en nuestra Evaluación integral de riesgos significativos para los derechos humanos (CSRA) de 2022, que publicamos en el [Informe sobre derechos humanos de 2022](#). En esta tabla, se brindan ejemplos ilustrativos de cómo abordamos los riesgos potenciales en 2025. Más adelante en este informe, analizaremos más en detalle algunos de estos ejemplos y cómo gestionamos posibles riesgos relacionados con la inteligencia artificial (IA), las elecciones y los conflictos.

1. Libertad de opinión y expresión

El [derecho a la libertad de opinión y expresión](#) incluye el derecho a buscar, recibir y compartir ideas e información de todo tipo. Es un derecho básico, fundamental para la protección de la dignidad humana, la autonomía individual y la democracia. La libertad de expresión es parte indispensable de nuestra misión, coherente con nuestro valor de brindar a todos la posibilidad de expresar su opinión.

Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA	Ejemplos de cómo Meta abordó los riesgos potenciales en 2025
Las políticas de moderación de contenido de Meta y su aplicación pueden limitar la libertad de expresión.	En enero de 2025, actualizamos la Política de conducta que incita al odio para incentivar el debate sobre algunos temas que forman parte del discurso general, como la inmigración y el género. Centramos nuestros sistemas automatizados en combatir las infracciones ilegales y de alta gravedad, como el terrorismo, la explotación sexual infantil, las drogas, el fraude y las estafas. Para las infracciones menos graves de las políticas, nos basamos en los reportes de los usuarios antes de tomar medidas. Además, ajustamos nuestros sistemas para que, en general, se requiera un nivel de certeza mucho mayor antes de que se elimine un contenido. Más adelante, en 2025, publicamos nuestro progreso en la precisión de la aplicación de las Normas comunitarias en nuestro Informe de cumplimiento de las Normas comunitarias, como parte de los Informes de integridad. Dejamos de aplicar los cambios que hicimos en 2021 para reducir la cantidad de contenido cívico que ven las personas y empezamos a recomendar más contenido político con base en señales personalizadas; por ejemplo, cuando un usuario indica que le gusta un contenido.

³ Para conocer la diferencia entre riesgo "inherente" y riesgo "residual", consulte el [Informe sobre derechos humanos de Meta de 2024](#), página 11.



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

Las políticas de moderación de contenido de Meta y su aplicación pueden limitar la libertad de expresión (continuación).

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

Concluimos nuestro programa de verificación de datos independiente en EE. UU. y, en su lugar, empezamos a implementar un programa de [notas de la comunidad](#). En el resto del mundo, seguimos basándonos en [verificadores de datos](#) que son independientes de Meta y están certificados por [International Fact-Checking Network](#), una red de organismos imparciales, o por [European Fact-Checking Standards Network](#), en Europa, para abordar la información errónea en Facebook, Instagram y Threads.

→ [Leer más](#)

El Gobierno impone límites excesivos sobre el contenido.

Nos esforzamos por cumplir los compromisos que asumimos con [Global Network Initiative](#) (GNI). En 2025, Meta se sometió a su cuarta evaluación de GNI (que abarcó el período de abril de 2022 a diciembre de 2024). El Consejo de GNI determinó de manera unánime que Meta había "hecho esfuerzos de buena fe para implementar los Principios de GNI, con una mejora a lo largo del tiempo". También informamos sobre nuestras [respuestas](#) a las solicitudes gubernamentales de datos o restricciones de contenido ([aquí](#) y [aquí](#)). En 2025, entre los [casos de éxito](#) destacados en materia de transparencia, relacionados con el discurso político, estuvieron los de Alemania, Austria, Colombia, Costa de Marfil, India, Irak, Italia, Kazajistán, Malasia, Nueva Zelanda, Polonia, Rumania, Singapur, Sudáfrica, Tanzania, Turquía y Vietnam.



Lanzamiento de las notas de la comunidad

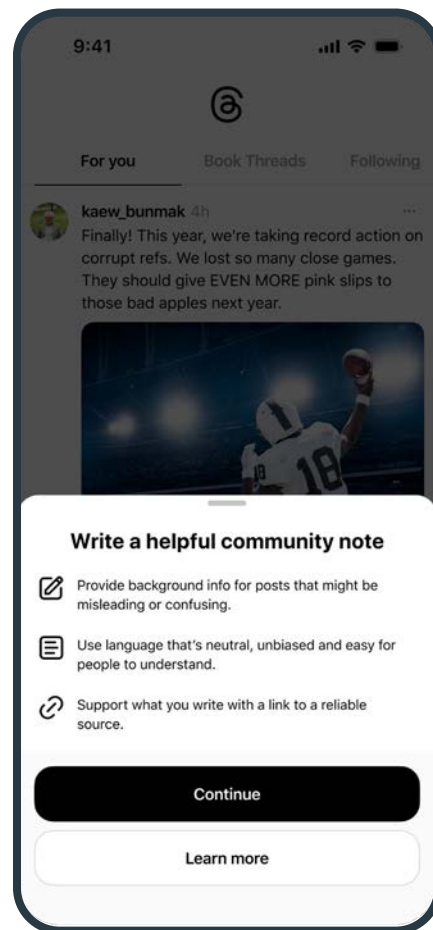
En 2025, Meta lanzó las [notas de la comunidad](#) en EE. UU. Las notas de la comunidad son una nueva herramienta para que la comunidad agregue contexto a contenido que pueda resultar engañoso o confuso. Nuestro enfoque se basó en los principios de derechos humanos, incluida la libertad de expresión. El objetivo era mejorar el acceso a información que goza de amplia confianza entre nuestra comunidad de usuarios, mediante intervenciones proporcionales que protejan la libertad de expresión.

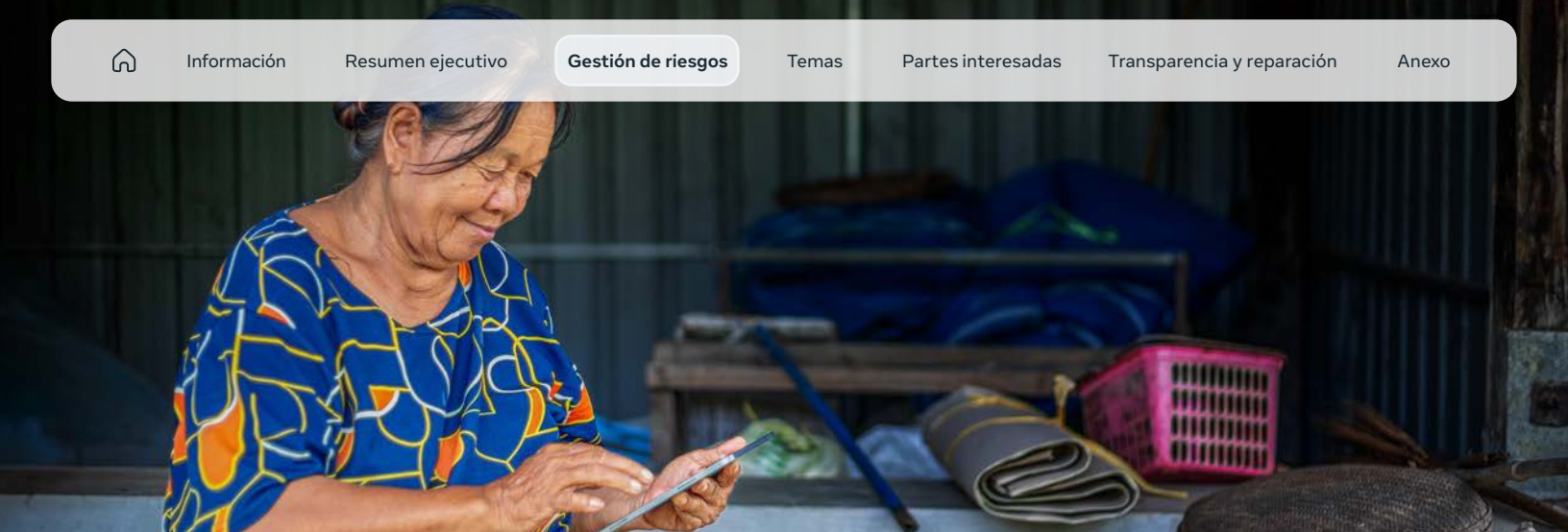
Los usuarios elegibles (mayores de 18 años con una cuenta verificada y en regla) escriben y califican las notas. Una nota solo se muestra cuando alcanza un umbral de "consenso útil". Esto significa que un número suficiente de colaboradores que antes no estaban de acuerdo ahora sí coinciden en que la nota es útil. Los nombres de los autores no se revelan para reducir la presión social y los riesgos de represalias, además de fomentar la participación, incluso en temas delicados.

Las notas de la comunidad pueden promover la libertad de expresión al agregar contexto, en lugar de eliminar o restar importancia a los mensajes legales, y al ampliar la participación más allá de un pequeño grupo de verificadores de datos profesionales. Además, están diseñadas con medidas de seguridad para ayudar a prevenir el uso indebido. Entre ellas, están los requisitos de integridad de la cuenta, las medidas de protección de URL y contra el spam, y el cumplimiento de las Normas comunitarias en las notas.

Meta pidió orientación al Consejo asesor de contenido sobre los factores nacionales que deberíamos tomar en cuenta al decidir si un país debe quedar excluido de la ampliación de nuestras notas de la comunidad, por ejemplo, los bajos niveles de libertad de expresión, la falta de libertad de prensa y los niveles de alfabetización digital.

En noviembre de 2025, el Consejo asesor de contenido aceptó la [solicitud de Meta de una opinión de asesoramiento normativo](#) para ayudarnos a definir el enfoque respecto a la expansión del programa de notas de la comunidad fuera de EE. UU. El Consejo asesor de contenido publicó su [asesoramiento](#) en marzo de 2026.





2. Privacidad

El [derecho a la privacidad](#) es condición necesaria para que se cumplan los demás derechos humanos, como la libertad de expresión, la libertad de reunión y asociación, y la libertad de culto y religión. Uno de los principios fundamentales que se detallan en nuestra [Política corporativa de derechos humanos](#) es mantener a las personas seguras y proteger la privacidad.

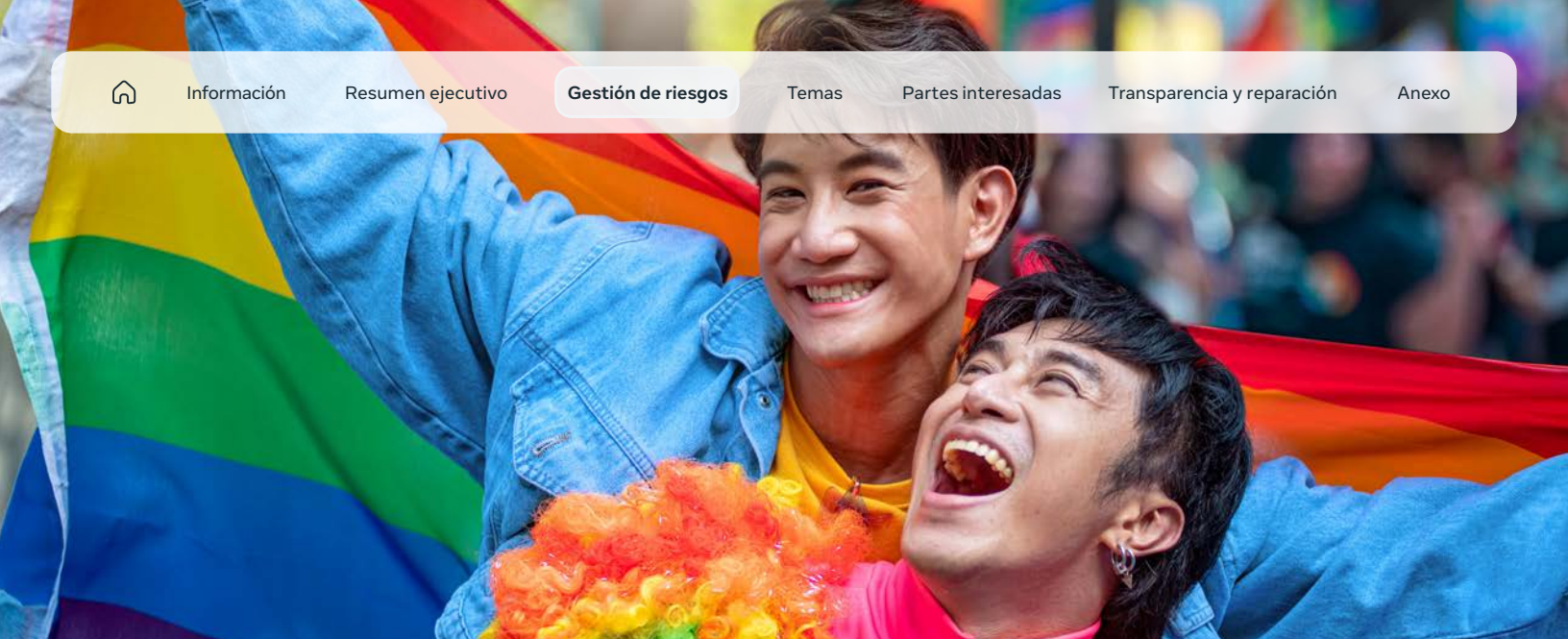
Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA

El contenido o los comportamientos en las apps de Meta pueden afectar de manera adversa los derechos de privacidad y protección de datos.

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

En 2025, desarrollamos el [tratamiento privado para Meta AI en WhatsApp](#), que está integrado en un entorno de ejecución confiable. Con el tratamiento privado, los usuarios pueden indicarle a Meta AI que trate sus solicitudes en un entorno de nube seguro y privado, donde nadie, ni siquiera Meta ni WhatsApp, pueda acceder a sus mensajes personales. Esta tecnología ahora sustenta ciertas funciones opcionales de Meta AI en WhatsApp, como resumir mensajes y ofrecer ayuda para redactar. Una vez que el tratamiento privado haya terminado de responder a una solicitud, los mensajes no se guardarán. Lea más sobre este trabajo en nuestra [Actualización de los avances en materia de privacidad](#).

Presentamos una [demanda](#) contra la entidad responsable de las apps CrushAI o de "nudificación", que permiten crear imágenes de desnudos o sexualmente explícitas generadas con IA de personas sin su consentimiento. Además, [anunciamos](#) las medidas adicionales que tomamos para poner freno a estas apps. Este trabajo complementó nuestras iniciativas de larga data para ayudar a prevenir que las imágenes íntimas abusivas, ya sean reales o generadas por IA, se difundan online, a través de herramientas como [StopNCII.org](#) y [Take It Down del Centro Nacional para Niños Desaparecidos y Explotados](#).



3. Igualdad y no discriminación

El [derecho a la igualdad y la no discriminación](#) brinda igualdad de protección contra la discriminación. Para honrar este derecho, no permitimos conductas de odio en nuestras plataformas, conforme se definen en nuestra [política](#).

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

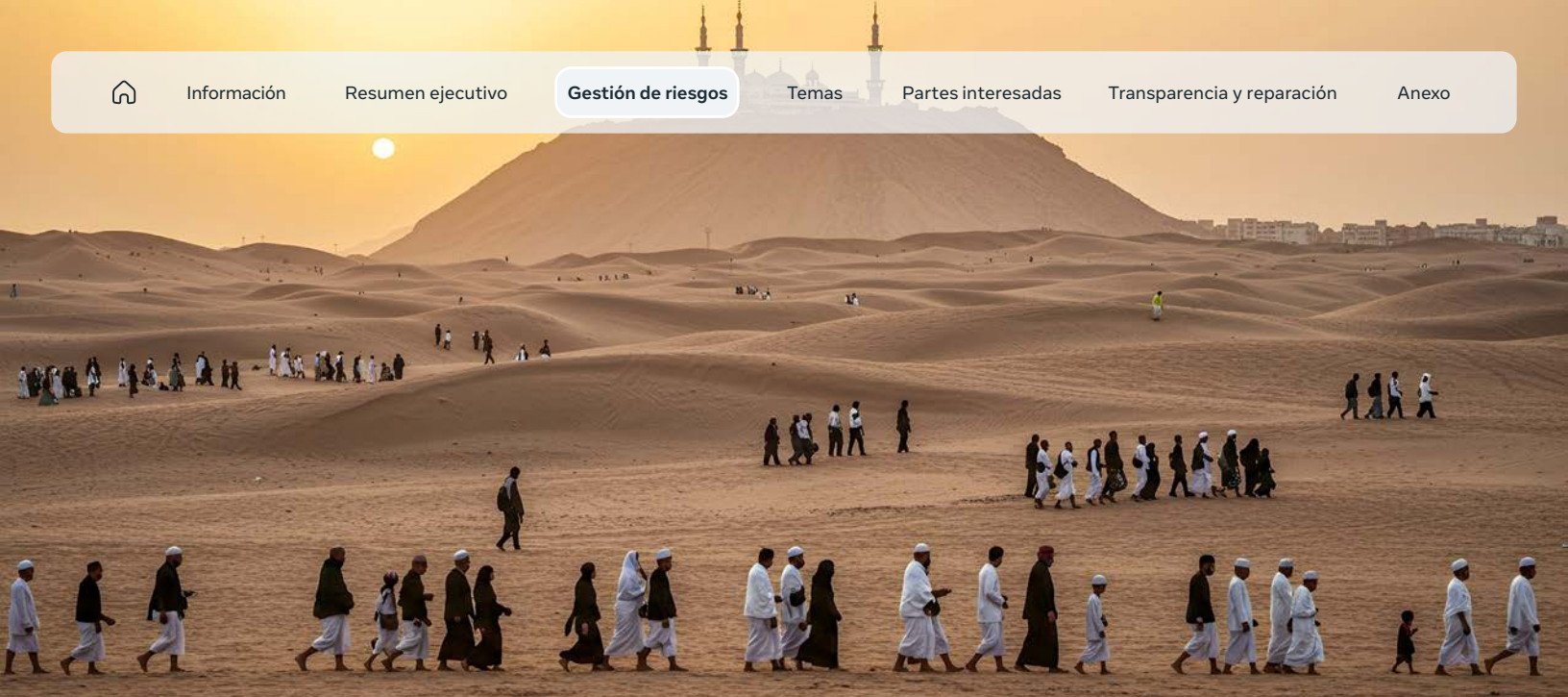
La moderación en algunos idiomas y dialectos puede suponer un mayor desafío.

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

[Presentamos](#) y publicamos en open source [Meta Omnilingual Automatic Speech Recognition](#), un conjunto de modelos que ofrecen funcionalidades de reconocimiento automático de voz (ASR) para más de 1.600 idiomas, incluidos 500 idiomas con pocos recursos. Al permitir que el contenido hablado se transcriba en idiomas que anteriormente no contaban con este servicio, esta tecnología abre nuevas oportunidades educativas. Asimismo, refuerza aún más la libertad de expresión, ya que brinda a las personas la posibilidad de acceder al conocimiento y compartir contenido en el idioma que mejor conocen.

Contenido que afecta de manera adversa la igualdad y la no discriminación (por ejemplo, la conducta que incita al odio).

Antes de los [cambios en las políticas y los productos](#) anunciados en enero de 2025, evaluamos sus riesgos y tomamos en cuenta las medidas de mitigación pertinentes, como la disponibilidad de otras políticas y los reportes de los usuarios, para abordar este tipo de contenido. En los países que atraviesan conflictos o crisis, es posible que apliquemos la detección proactiva de forma predeterminada en algunos ámbitos de las políticas, como las infracciones relacionadas con la conducta que incita al odio. Por ejemplo, usamos la detección proactiva en Libia después de la implementación del Protocolo de la política de crisis.



4. Vida, libertad y seguridad de la persona

El [derecho a la vida, la libertad y la seguridad de la persona](#) concierne a la libertad de daños físicos y reclusión. Para Meta, respetar este derecho humano incluye mitigar el riesgo de que el contenido provoque daños, incluidos riesgos de violencia y trata de personas, amenazas online amparadas por el Estado y la participación de grupos no pertenecientes al Estado en hechos de odio o violencia o el apoyo a estos.

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

- Personas malintencionadas que:
- Explotan las apps y los servicios de Meta para coordinar daños online y en la vida real.
 - Usan indebidamente apps y servicios para realizar ciberataques o phishing.
 - Amenazan y acosan a defensores de los derechos humanos, activistas y otros grupos vulnerables.

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

Tras una investigación exhaustiva y consultas con sus socios, Meta identificó un posible aumento del riesgo de reclutamiento de niños soldados en Sudán, por lo que elaboró mensajes sobre seguridad para los jóvenes en árabe, francés e inglés, adaptados a la región. En noviembre de 2025, Meta lanzó dos campañas complementarias: una promoción rápida dirigida a aproximadamente un millón de jóvenes sudaneses (de 13 a 18 años) en Facebook e Instagram, y un mensaje emergente interactivo y personalizado enviado a unos 40.000 jóvenes conectados a las [desarticulaciones estratégicas de redes](#). Los mensajes advertían sobre los riesgos del reclutamiento y les indicaban a los jóvenes dónde encontrar recursos externos de seguridad.

[→ Leer más](#)

Seguimos respaldando el Fondo para defensores de derechos humanos y nuestro [Programa de socios de confianza](#) a fin de mejorar la respuesta ante emergencias en favor de los defensores de los derechos humanos y otros grupos vulnerables.

5. Interés superior del niño

La Convención sobre los Derechos del Niño de la ONU (UNCRC) afirma que, en todas las medidas concernientes a los niños, "el interés superior del niño debe ser una consideración primordial". El [marco de protección de los intereses de los niños](#) de Meta concuerda con los valores fundamentales de la UNCRC. La protección de los niños online es una prioridad principal para Meta. Ofrecemos herramientas para adolescentes, padres y tutores con protecciones incorporadas para protegerlos, al tiempo que les brindamos un espacio para ejercer su derecho a la libertad de expresión y acceder a información.

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

Los niños pueden quedar expuestos a contenido no deseado e inapropiado o a comportamientos depredadores.

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

A partir de las [cuentas de adolescentes](#), agregamos [nuevas medidas de protección](#) y presentamos [configuraciones de contenido actualizadas](#), inspiradas en los criterios de clasificación de películas para mayores de 13 años y en los comentarios de los padres. Esto significa que el objetivo de Instagram es que, de forma predeterminada, los adolescentes vean contenido similar al que verían en películas adecuadas para su edad. Además, presentamos una nueva configuración más estricta destinada a los padres que prefieren una experiencia más restrictiva para sus hijos adolescentes. También anunciamos nuevas medidas para ayudar a los padres a controlar cómo sus [hijos adolescentes interactúan con la IA](#) en nuestras apps, [tomamos medidas legales](#) contra las empresas responsables de las apps de "nudificación" y [nos asociamos con escuelas para combatir el ciberbullying](#).

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

El [derecho a la participación pública, al voto y a presentarse a una candidatura](#) en elecciones libres y equitativas es un pilar de la democracia. Proteger la integridad de las elecciones en nuestras apps y servicios es una de nuestras principales prioridades. Trabajamos arduamente para proteger las elecciones online antes de los períodos electorales, en el transcurso de estos y una vez que finalizan.

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

El contenido infractor puede repercutir de manera negativa en la participación pública, el voto o la candidatura a un cargo público. Esto puede surgir de actividades, por ejemplo:

- Agentes malintencionados coordinados que interfieren en procesos electorales
- Amenazas de daño y violencia en la vida real contra candidatos
- Iniciativas individuales para impedir que las personas voten, aumento de la cantidad de spam, injerencia extranjera o reportes de contenido que infringe nuestras políticas

Ejemplos de cómo Meta abordó los riesgos potenciales en 2025

Continuamos con las iniciativas electorales, que se apoyan en nuestro enfoque electoral a escala y basado en el riesgo, como se detalla en el [Informe sobre derechos humanos de 2024](#). En 2025, nuestros preparativos electorales incluyeron los de Alemania, Australia, Camerún, Chile, Costa de Marfil, Ecuador, Egipto, Irak, Japón, Myanmar, Filipinas, Polonia, Rumania y Tanzania, entre otros países.

7. Libertad de reunión y asociación

El [derecho a la libertad de reunión y asociación](#) es fundamental para la democracia y es interdependiente de muchos otros derechos que garantizan las leyes internacionales de derechos humanos, incluido el derecho a la libertad de expresión y a formar parte de asuntos públicos. Para Meta, este derecho se relaciona con nuestros valores fundamentales de brindar a las personas una voz y forjar conexiones y comunidad.

Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA	Ejemplos de cómo Meta abordó los riesgos potenciales en 2025
Ciertos tipos de contenido en nuestras plataformas pueden hacer que algunas personas se sientan incapaces de conectarse libremente en las apps de Meta u offline.	Implementamos el Protocolo de la política de crisis a fin de dotar de recursos a nuestras iniciativas para abordar el contenido infractor relacionado con manifestaciones masivas, por ejemplo, las "protestas de la generación Z" en Irán, Madagascar, Nepal y Serbia.

8. Derecho a la salud

El [derecho a la salud](#) es el derecho que todos tenemos a acceder al más alto estándar posible de salud física y mental. Para honrar este derecho, Meta aumenta el acceso a información de salud, permite que personas con problemas de salud similares se conecten entre sí y les brinda herramientas para tomar decisiones fundamentadas sobre su salud y bienestar.

Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA	Ejemplos de cómo Meta abordó los riesgos potenciales en 2025
Contenido que infringe las normas que incita a cometer daños en la vida real o se diseñó con ese fin.	Anunciamos medidas de protección para los adolescentes en nuestros servicios de IA; por ejemplo, diseñarlos para que respondan de manera segura a las consultas sobre autolesiones, suicidio y trastornos alimentarios. También anunciamos nuevas funciones de control, como el lanzamiento de estadísticas para padres sobre los temas de los que sus hijos adolescentes hablan con la IA. En EE. UU., nos propusimos aumentar nuestras iniciativas para combatir el bullying online en las escuelas, brindándoles a los maestros, educadores y administradores un sistema priorizado para reportar casos relacionados con la seguridad de los adolescentes.



Temas destacados

Innovar respetando los derechos humanos

Las innovaciones en tecnología de dispositivos portátiles e IA brindaron oportunidades importantes para promover los derechos humanos y respaldar la mitigación responsable de riesgos.

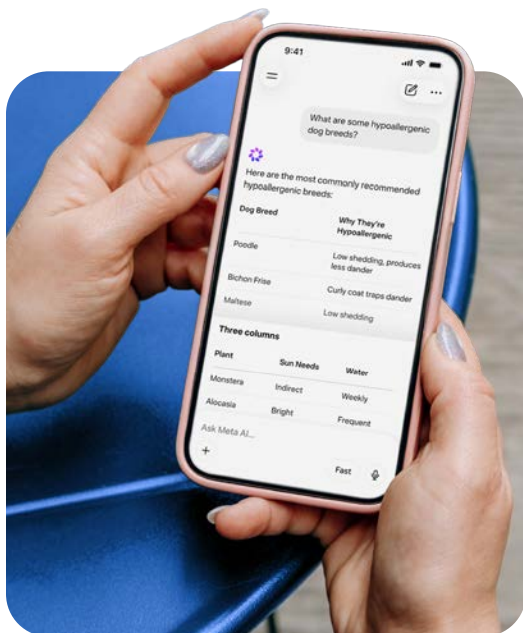
Mejorar la accesibilidad con la tecnología de dispositivos portátiles

En 2025, [lanzamos Meta Ray-Ban Display y Meta Neural Band](#), una combinación de lentes con IA de nueva generación y una pulsera que usa la tecnología de [electromiografía de superficie \(sEMG\)](#). Con esta tecnología, las señales eléctricas que generan los movimientos sutiles de los músculos se convierten en comandos que permiten controlar el dispositivo sin usar las manos.

Los prototipos de la pulsera se desarrollaron pensando en [la accesibilidad](#) y respetando el derecho a la igualdad y a la no discriminación. Con base en un estudio en el que participaron casi 200.000 personas, esta pulsera ofrece [nuevas posibilidades](#) a quienes tienen limitaciones de movilidad por lesiones en la médula espinal, accidentes cerebrovasculares u otras condiciones discapacitantes, así como a quienes sufren temblores o tienen menos de cinco dedos. Para proteger la privacidad, el aprendizaje automático que convierte las señales de EMG en comandos digitales se [procesa localmente en el dispositivo](#).



Junto a los lentes con IA, nuestra tecnología [ayuda a las personas con discapacidad](#) a desenvolverse en la vida diaria con mayor independencia. Funciones como las llamadas con manos libres, la traducción en tiempo real y los subtítulos instantáneos eliminan las barreras de comunicación. Asimismo, gracias a la función [Llama a un voluntario](#) (creada en colaboración con [Be My Eyes](#)), las personas ciegas o con baja visión pueden conectarse con voluntarios para recibir descripciones en vivo de lo que ven a su alrededor. Al desarrollar estas herramientas, colaboramos directamente con organizaciones de personas con discapacidad, como la [Asociación de Veteranos Ciegos](#) y los [Centros de Rehabilitación para Ciegos](#) del Departamento de Asuntos de los Veteranos de EE. UU., para elaborar guías de capacitación y apoyar casos de uso relacionados con la accesibilidad.



Seguir innovando de manera responsable en el campo de la IA

Lanzamos el [asistente de Meta AI independiente](#) junto con una versión actualizada de un macromodelo lingüístico en open source, [Llama 4](#). También presentamos la nueva tecnología de [tratamiento privado, que utiliza un entorno de ejecución confiable](#). Con ella, es posible habilitar el asistente de Meta AI en WhatsApp sin dejar de lado las medidas de protección de la privacidad, como las comunicaciones cifradas. Para obtener más detalles, consulte nuestro informe anterior sobre diligencia debida, en el que se abordan los [beneficios para los derechos humanos del cifrado de extremo a extremo](#).

Para reducir los riesgos de uso indebido y las respuestas potencialmente dañinas de nuestros productos y modelos de IA, el desarrollo de Llama 4 incluyó el filtrado de datos durante el preentrenamiento, medidas de mitigación posteriores al entrenamiento y medidas de mitigación en el nivel del sistema. Estos se explican en nuestra [guía de uso para desarrolladores sobre medidas de protección de IA](#).

También publicamos nuestro [marco de IA de vanguardia](#), que se enfoca en los riesgos más críticos en las áreas de amenazas a la ciberseguridad y riesgos derivados de armas químicas y biológicas. En este marco, describimos varios procesos que seguimos para anticipar y mitigar los riesgos al desarrollar sistemas de IA de vanguardia, entre ellos:

Identificar consecuencias catastróficas para prevenirlas



Nuestro marco identificó posibles consecuencias catastróficas relacionadas con riesgos cibernéticos, químicos y biológicos que nos esforzamos por prevenir. Se centró en evaluar si los avances tecnológicos habían propiciado esos resultados catastróficos y, de ser así, identificar formas de mitigar tales riesgos.

Ejercicios de modelado de amenazas



Llevamos a cabo ejercicios de modelado de amenazas para anticipar cómo distintos actores podrían intentar hacer un mal uso de la IA de vanguardia, con el fin de provocar esos resultados catastróficos, trabajando con especialistas externos cuando fue necesario. Estos ejercicios resultaron clave para nuestro enfoque centrado en los resultados.

Establecer umbrales de riesgo

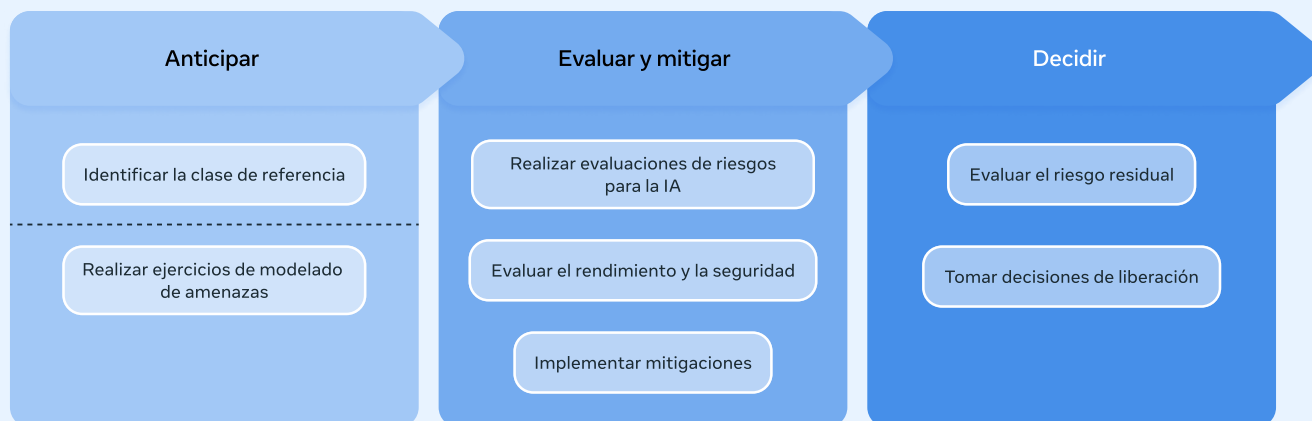


Definimos los umbrales de riesgo en función del grado en que nuestros modelos podían simular las situaciones de amenaza. Contamos con procesos para mantener los riesgos dentro de niveles aceptables, lo que incluye la aplicación de medidas de mitigación.

Además, seguimos respaldando un amplio ecosistema para el desarrollo y la implementación responsables de la IA. Para que los desarrolladores puedan crear aplicaciones seguras, ampliamos nuestra inversión en [medidas de seguridad en el nivel del sistema](#). Entre ellas, están [Llama Guard 4](#), que ofrece medidas de seguridad unificadas para modalidades de texto e imagen, y [Llama Prompt Guard 2](#), que mejora la detección de fugas de seguridad y la inyección de indicaciones.

Nuestro trabajo se basó en una amplia serie de interacciones y consultas con la sociedad civil, las organizaciones internacionales, el mundo académico y el público en general sobre los aspectos relacionados con los derechos humanos de la IA. Esto incluyó [Foros de la comunidad](#) (organizados en colaboración con el Laboratorio de Democracia Deliberativa de la Universidad de Stanford), en los que se reunieron diversas perspectivas sobre el futuro de los agentes de IA y su gobernanza.

Marco de la IA de frontera 2025





Herramientas de bienestar en open source para la revisión de contenido

A fin de que los revisores apliquen nuestras políticas de manera precisa, también deben enfocarse en su [resiliencia y bienestar](#).

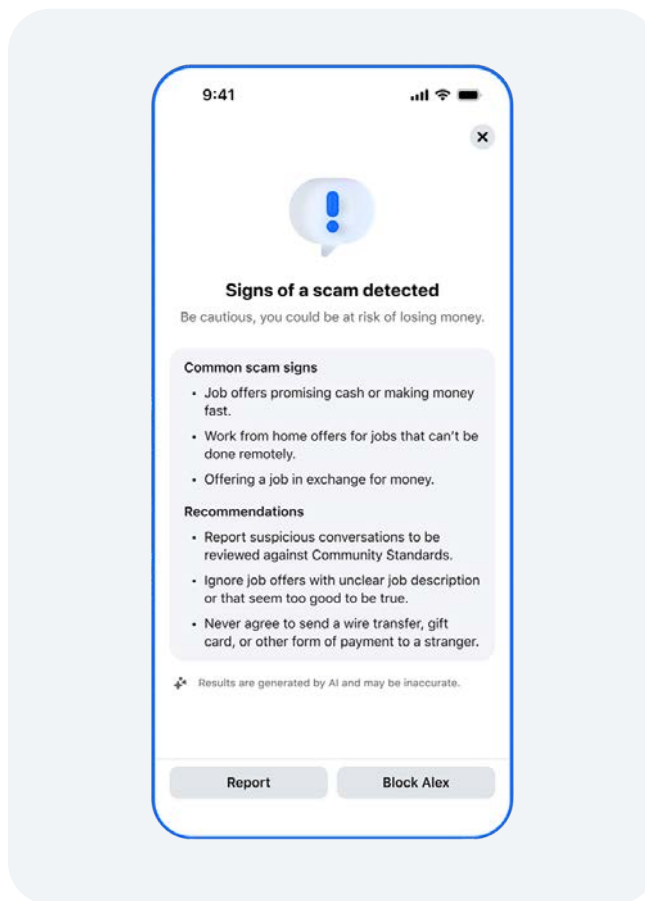
En 2025, [pusimos a disposición en open source un conjunto de filtros de revisión de contenido](#) diseñados para apoyar el bienestar de los moderadores. De esta manera, otros desarrolladores pueden aprovechar nuestra experiencia en este ámbito. Estos filtros personalizables para imágenes y videos (con funciones como desenfoque, conversión a escala de grises o sepia, reducción de detalles y pantallas de advertencia predictivas con IA) pueden ayudar a reducir los posibles efectos negativos en la salud mental que puede tener la exposición a contenido explícito o perturbador.

Fraudes y estafas

Las redes de fraudes y estafas crecieron mucho en la última década y afectaron de manera negativa los derechos humanos de las personas. Muchos centros de estafas se valen de la trata de personas y el trabajo forzado para llevar a cabo sus actividades, y las víctimas de estas estafas pueden sufrir pérdidas económicas considerables y otros daños. En 2025, [eliminamos](#) más de 159 millones de anuncios fraudulentos por infringir nuestras políticas. Además, [cerramos](#) 10,9 millones de cuentas en Facebook e Instagram relacionadas con redes criminales de estafadores.

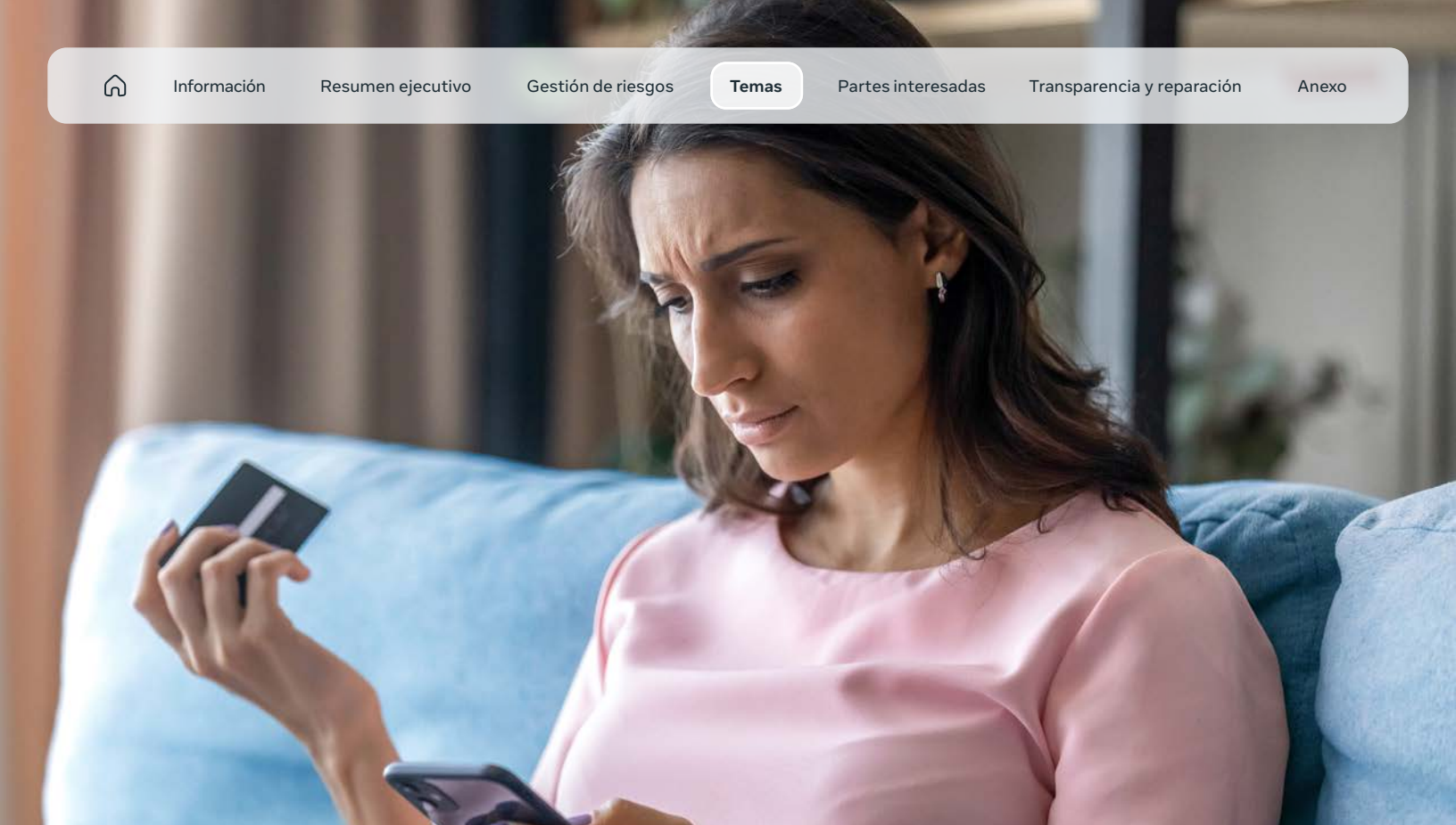
Dimos prioridad a la seguridad mediante [medidas de protección de la plataforma](#), como [el reconocimiento facial](#) para combatir la suplantación de identidad, [las advertencias de detección de estafas](#) en Messenger y [las advertencias sobre el uso compartido de pantalla](#) en WhatsApp. Nuestra [labor](#) para desarticular a los estafadores también incluyó litigios y el apoyo a las autoridades a fin de hacer que rindan cuentas. Lograr un equilibrio entre la seguridad y el bienestar de nuestras comunidades y las medidas de mitigación que sean proporcionales y respeten los derechos puede ser un desafío. Tomamos estas medidas pensando en la privacidad y de acuerdo con las normas internacionales de derechos humanos, como los [Principios de Global Network Initiative](#).

Los fraudes y las estafas son desafíos en todo el mundo. Para que el trabajo sea efectivo, es necesario compartir señales, colaborar entre las comunidades de defensores y adoptar un enfoque que involucre a toda la sociedad. A lo largo de 2025, Meta participó en asociaciones para compartir señales, como [Global Signal Exchange](#), [Global Anti-Scam Alliance](#) y [Fraud Intelligence Reciprocal Exchange](#). También [abogamos](#) por que se publiquen listas de los países que no toman las medidas adecuadas y por que se impongan sanciones internacionales. Todo esto sin dejar de lado los derechos humanos y las consecuencias no deseadas para las poblaciones vulnerables, como las personas víctimas de trata a las que se obliga a participar en operaciones de estafa.



Desarticular redes criminales de estafas en el Sudeste Asiático

En la última década, [surgieron](#) sofisticadas redes y complejos de estafas a escala industrial. En toda [la región de Asia-Pacífico](#), estos complejos dan trabajo a cientos de miles de personas y operan casi con total impunidad. Los centros de estafas criminales atraen a las personas que buscan empleo con publicaciones engañosas y las obligan a trabajar como estafadores online. A menudo, lo hacen bajo la amenaza de abuso físico, lo que viola la prohibición de la esclavitud, la servidumbre y el trabajo forzado, conforme a las normas internacionales de derechos humanos.



Entre enero y octubre de 2025, [eliminamos](#) más de 6.400 cuentas y páginas de Facebook por [infringir](#) nuestras políticas contra [fraudes, estafas y prácticas engañosas](#) y [personas y organizaciones peligrosas](#). Una red desarticulada con origen en Camboya [se hizo pasar](#) por autoridades policiales en todo el mundo para ofrecer falsamente servicios de recuperación de pérdidas. Este trabajo requirió una colaboración transfronteriza y la estrecha supervisión del impacto de cambios geopolíticos, como el conflicto en Myanmar y a lo largo de la frontera en el norte de Camboya. Tratamos estas iniciativas más detalladamente en el [Informe de amenazas adversas](#) del primer semestre de 2026.



Las iniciativas de lucha contra la delincuencia en 2025 culminaron con nuestra participación en un programa piloto llamado [Semana Conjunta contra la Delincuencia](#) en Bangkok, donde Meta colaboró con el FBI, la Policía Real de Tailandia y otras agencias de seguridad pública. Este programa piloto [permitió](#) la eliminación de 59.000 cuentas, páginas y grupos de las plataformas de Meta, así como seis órdenes de arresto.



Además de las desarticulaciones, actualizamos nuestras [Normas comunitarias sobre explotación de personas](#) para detectar más contenido de reclutamiento de posibles redes criminales de estafas. La actualización limita las publicaciones de ofertas de trabajo en lugares con alto riesgo de explotación laboral, según lo confirmen las autoridades y las organizaciones no gubernamentales locales.

Meta también se asoció con la Oficina de las Naciones Unidas contra la Droga y el Delito en la campaña [#TrappedInScamCrime](#) para ayudar a combatir el reclutamiento online y la trata de personas con fines de actividades delictivas forzadas, vinculadas a los centros de estafas del Sudeste Asiático. La campaña difundió mensajes de prevención basados en las experiencias de los sobrevivientes para ayudar a las personas a reconocer las ofertas de trabajo engañosas y acceder a recursos de apoyo.

Asimismo, colaboramos con la Asociación de Naciones del Sudeste Asiático a fin de desarrollar [principios](#) para combatir las estafas online. Este trabajo se sumó a nuestras [iniciativas más amplias contra la trata de personas](#).

Cómo nos preparamos para afrontar una crisis y respondemos ante estas situaciones

Nos preparamos para afrontar muchas crisis, incluidos conflictos, violencia intracomunitaria, disturbios civiles, manifestaciones masivas y desastres ambientales, así como ataques terroristas y tiroteos, en todo el mundo. En 2025, pusimos en marcha y coordinamos respuestas ante situaciones de crisis en Camerún, India (Cachemira), Indonesia, Libia, Madagascar, Nepal, Serbia, Tanzania y Tailandia-Camboya, entre otros países y territorios. Mantuvimos nuestras iniciativas respecto de crisis designadas en virtud del [Protocolo de la política de crisis](#) en los conflictos acontecidos en Ucrania, Sudán y Oriente Medio. Abordamos estas crisis específicas en nuestros [informes anteriores sobre derechos humanos](#).

En períodos de crisis, nos apoyamos en el Protocolo de la política de crisis como herramienta clave. El protocolo nos guía en el uso rápido de los mecanismos para mitigar posibles daños, tal como se describe en el [Informe sobre derechos humanos de 2024](#). Los tipos específicos de respuestas implementados son coherentes con los riesgos observados. Se basan en intervenciones en crisis anteriores, principios de derechos humanos y el derecho internacional humanitario.

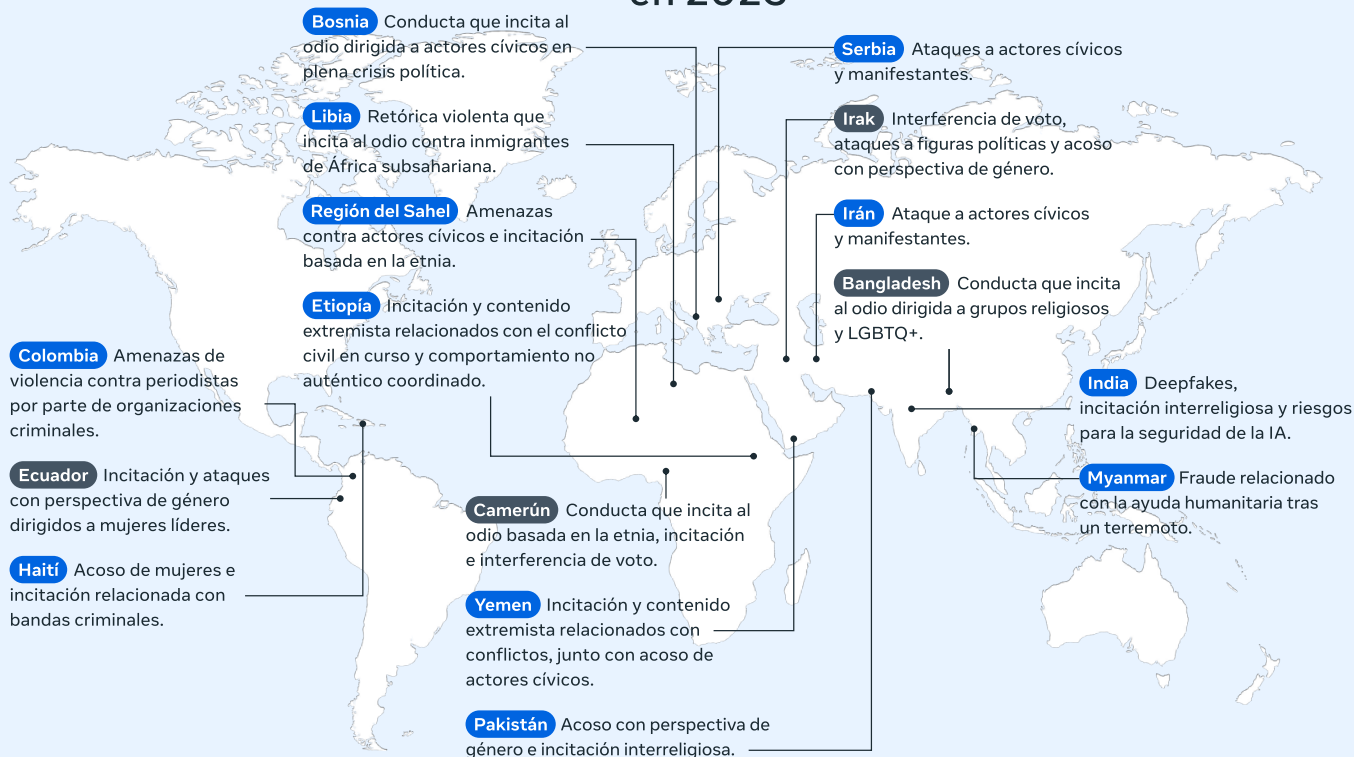
En la página siguiente, mostramos cómo nos preparamos y respondemos ante crisis y en zonas afectadas por conflictos y de alto riesgo, usando un ejemplo de una región que no apareció en nuestros informes anteriormente.

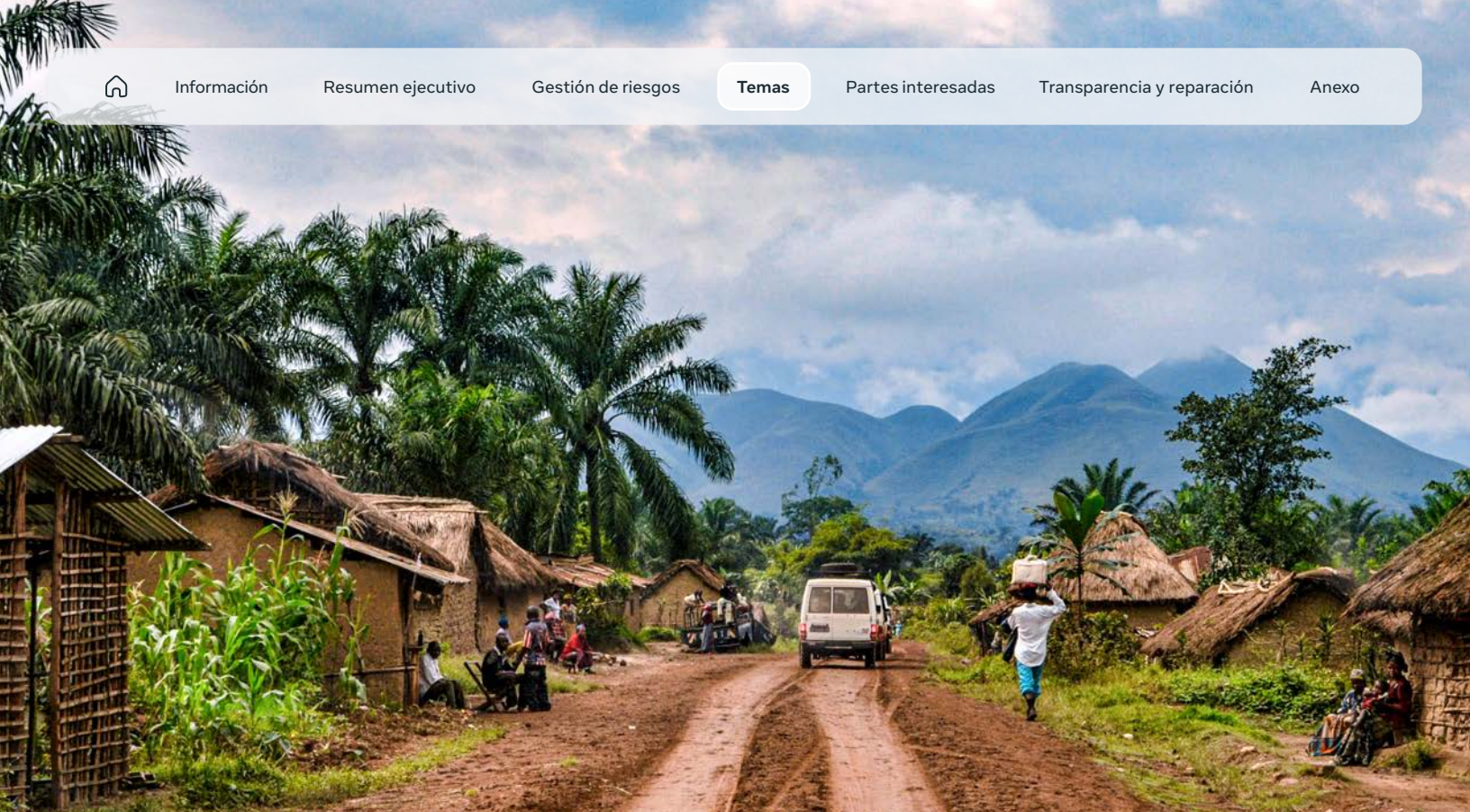
Leyenda

Respuesta ante emergencias

Elecciones

Aspectos destacados de la respuesta ante emergencias y la integridad de las elecciones en 2025





República Democrática del Congo

En 2025, los combates en el este de la República Democrática del Congo (RDC) se intensificaron cuando el grupo armado M23, que habría recibido [respaldo](#) de Ruanda, se apoderó de territorio en Kivu del Norte. Esto provocó desplazamientos masivos de la población civil y [graves violaciones de los derechos humanos](#). Nuestros socios de confianza nos alertaron sobre el escalamiento de la violencia en el este del país y el riesgo creciente de violencia entre comunidades. Por consiguiente, Meta decidió clasificar esta situación como una crisis según el Protocolo de la política de crisis.

Un equipo multidisciplinario de especialistas de Meta, que abarca las áreas de políticas, operaciones, seguridad y políticas públicas, había estado supervisando la situación offline en la RDC desde que el conflicto se reavivó a finales de 2021. Para ello, tomó medidas contra el contenido que infringía nuestras Políticas de [información errónea y daño](#), [violencia e incitación](#), [organización de actos dañinos y promoción de la delincuencia](#) y [personas y organizaciones peligrosas](#). También habíamos realizado la implementación de medidas de mitigación para abordar posibles riesgos relacionados con el contenido online. Entre ellas, la identificación de afirmaciones potencialmente dañinas que ya se habían revisado con antelación, de acuerdo con la Política de información errónea y daño, y el uso de la [Política de lugares de alto riesgo temporal](#). Además, anteriormente habíamos identificado grupos de riesgo relacionados con "ruandófonos, ruandeses, banyamulenges o tutsis" para ayudar a prevenir el riesgo de sufrir daños en la vida real.

A medida que el escalamiento del conflicto se producía en 2025, evaluamos y mitigamos los riesgos relacionados con el contenido. Brindamos aclaraciones y orientación sobre nuestras políticas, revisamos las palabras clave para garantizar una detección eficaz y un cumplimiento de normas preciso, y aplicamos la Política de lugares de alto riesgo temporal.

Meta también recibió asesoramiento de socios de confianza para ayudar a definir la posible designación de nuevas afirmaciones dañinas previamente revisadas (declaraciones falsas o engañosas que podrían contribuir directamente a un riesgo de daño físico inminente, según nuestra Política de información errónea y daño). Gracias a la información que nos brindaron nuestros socios de confianza, pudimos detectar y eliminar de manera más efectiva el contenido que infringía las políticas, por ejemplo, el material que pudiera avivar las divisiones locales.

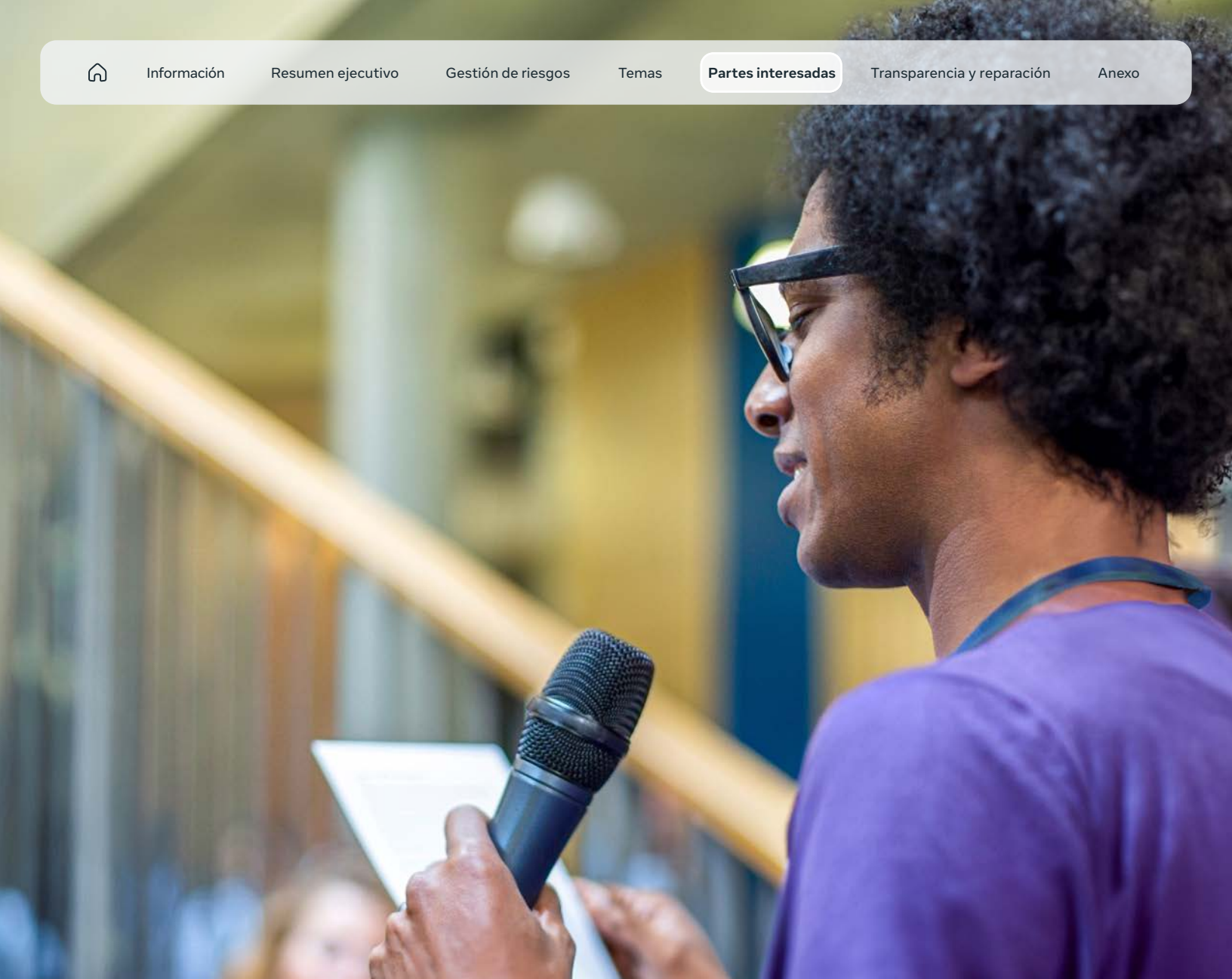
Participación de partes interesadas

La [colaboración](#) proactiva y estructurada con nuestra comunidad mundial de usuarios permite forjar las políticas de Meta y es el pilar de nuestras medidas para abordar los riesgos que ponen en peligro los derechos humanos.

En 2025, colaboramos con una amplia gama de partes, incluidos miembros de la sociedad civil, académicos, laboratorios de ideas, especialistas en derechos humanos y autoridades reguladoras. Entre los temas clave de políticas se incluyeron nuestro enfoque respecto a la inteligencia artificial (IA) y la moderación de contenido, así como el impacto de los cambios en los productos y las políticas presentados en enero de 2025, como las [notas de la comunidad](#), además del fraude y las estafas.

Colaboramos con más de 615 partes interesadas externas de 65 países para que sus aportes sirvieran de base en el desarrollo de políticas en varias áreas, como las personas y organizaciones peligrosas, así como las políticas de publicidad sobre salud y bienestar. Consultamos a cientos de personas más para las áreas de pronóstico de riesgos, análisis de tendencias, aplicación de la ley y prevención.

También colaboramos con la sociedad civil en varios temas relacionados con la libertad de expresión y la regulación, como en Argentina y Brasil, además de Malasia como parte de la [Conferencia de la Asamblea de Derechos Digitales de Asia-Pacífico](#). También impartimos capacitación sobre nuestras políticas para la sociedad civil. En África del Norte, por ejemplo, capacitamos a grupos de la sociedad civil sobre nuestras Políticas de [información errónea y daño](#) y de [conducta que incita al odio](#) para ayudarlos a identificar contenido que pudiera infringirlas. La capacitación también ayudó a que los defensores de los derechos humanos y los grupos vulnerables tuvieran herramientas para proteger su presencia digital.



Socios de confianza

Los socios de confianza son importantes aliados a la hora de identificar infracciones graves a las Normas comunitarias. Seguimos trabajando con [socios de confianza](#) para identificar tendencias, comprender mejor el impacto del contenido y el comportamiento online en las comunidades locales, y explorar cómo podemos fortalecer nuestros canales de escalamiento en la sociedad civil.

En 2025, los socios de confianza aportaron información valiosa y detectaron contenido perjudicial en países y territorios que sufrieron mayor agitación social o celebraron elecciones, por ejemplo, Afganistán, Camboya, Costa de Marfil, República Democrática del Congo, India, Indonesia, Israel, Kenia, Líbano, Madagascar, México, Nigeria, Palestina, Somalia, Sudán, Siria, Tailandia, Tanzania, Ucrania y Venezuela, entre otros países y regiones.

Eliminamos más de 60.000 contenidos infractores gracias al programa de socios de confianza. En la página siguiente, presentamos un ejemplo del impacto de los socios de confianza en Latinoamérica.

CASO DE ÉXITO

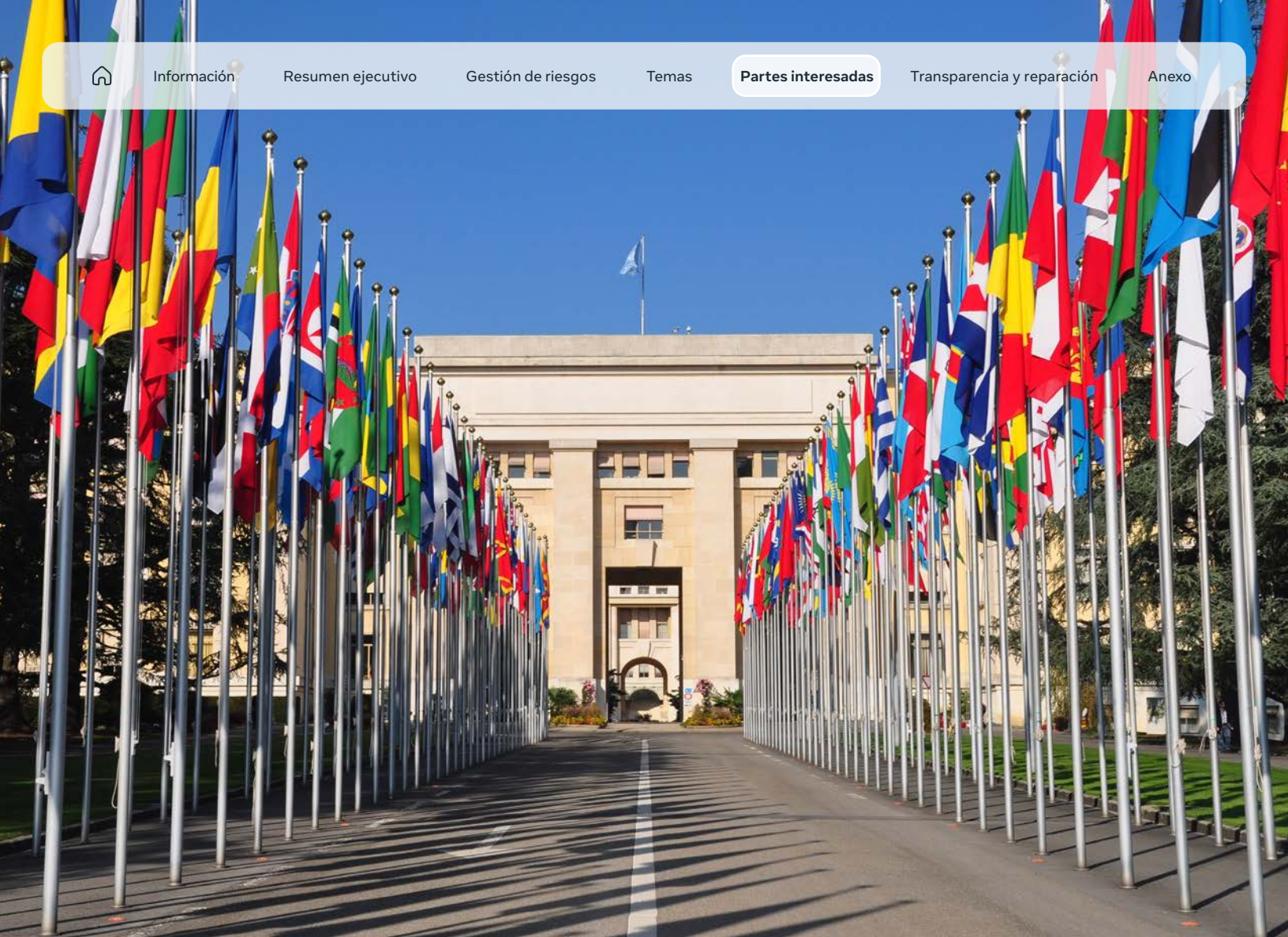
Información valiosa para prevenir el reclutamiento de jóvenes en Latinoamérica



En Latinoamérica, los socios de confianza desempeñaron un papel fundamental al reportar contenido que incluía la glorificación, el apoyo y la representación de [personas y organizaciones peligrosas](#). Además, aportaron comentarios sobre las iniciativas para prevenir el reclutamiento de jóvenes.

En Brasil, los socios de confianza reportaron contenido en Instagram que mostraba a miembros de una organización criminal designada, incluyendo imágenes de armas de fuego prohibidas, actos de violencia y drogas. En Colombia, los socios de confianza reportaron perfiles relacionados con grupos criminales, en el contexto de amenazas y ataques violentos por parte de estos grupos contra periodistas en el departamento de Guaviare. En Haití, el contenido reportado ensalzaba los actos delictivos de ciertas personas vinculadas a pandillas. Esta información valiosa respaldó nuestras iniciativas para eliminar el contenido que infringía las normas y reducir el riesgo que representan las personas y organizaciones peligrosas en nuestras plataformas, así como el de ataques físicos.

Meta se asoció con socios de confianza y especialistas regionales para lanzar campañas de comunicación centradas en los jóvenes en toda Latinoamérica. Estas campañas se dirigieron a 80.000 usuarios de entre 13 y 25 años en México y a otros 55.000 en Colombia, Haití y Brasil. Su objetivo era informar a los usuarios sobre las infracciones en la plataforma, educarlos sobre los riesgos de interactuar con personas y organizaciones peligrosas, y empoderarlos para que tomaran sus propias decisiones online con mayor conocimiento de causa. La campaña produjo un aumento medible en el número de desactivaciones de cuentas y de personas que se eliminaron de la lista de amigos.



Organizaciones internacionales

Meta siguió colaborando con el sistema de la ONU durante todo el año. Participamos en los debates sobre políticas de la 80.ª Asamblea General de las Naciones Unidas (UNGA). Entre los temas tratados, se incluyeron la gobernanza global de la tecnología digital y la IA, el uso de la IA para la [seguridad nacional](#) y cómo cerrar la brecha de capacidades en materia de IA. De igual manera, participamos en el lanzamiento del Diálogo Global de la ONU sobre la gobernanza de la IA, en el marco de la Semana de Alto Nivel de la UNGA.

Meta siguió trabajando en estrecha colaboración con la [Oficina del Alto Comisionado para los Derechos Humanos](#), incluido el [Proyecto B-Tech](#). Respondimos a solicitudes y nos asesoramos con la [Comisión Africana de Derechos Humanos](#) y de los Pueblos y los Procedimientos Especiales de Derechos Humanos de la ONU (especialistas independientes en derechos humanos), incluidos los Relatores Especiales sobre la libertad de opinión y de expresión. También colaboramos con otras agencias de la ONU y organizaciones humanitarias para orientar nuestro trabajo de respuesta a la crisis.

Además, participamos en el [Foro de Gobernanza de Internet](#) de 2025, el [Foro de las Naciones Unidas sobre Empresas y Derechos Humanos](#) y la [Cumbre RightsX](#), donde hablamos sobre casos de uso positivos de la IA para los defensores de los derechos humanos, entre otros temas. Seguimos participando y colaborando con las partes interesadas en otros foros de carácter multilateral, como el [Foro Global de Internet para Combatir el Terrorismo](#), [RightsCon](#) y [Tech Against Trafficking](#).

Transparencia y reparación



El [Consejo asesor de contenido](#), un organismo independiente, nos ayuda a resolver algunos de los asuntos más difíciles sobre la libertad de expresión online: qué contenido eliminar, cuál conservar y por qué. Este organismo revisa casos que remite Meta o que apelaron usuarios de Facebook, Instagram o Threads que están en desacuerdo con nuestras decisiones de moderación de contenido, y proporciona resoluciones vinculantes sobre si se debe eliminar o conservar contenido. El Consejo asesor de contenido también proporciona recomendaciones para mejorar nuestras prácticas de moderación de contenido y ofrece opiniones de asesoramiento normativo si así se le solicita.



En 2025, el Consejo asesor de contenido consideró casos respecto de las medidas de aplicación de políticas que tomamos sobre el contenido en vista del marco internacional de derechos humanos, como la libertad de expresión, los derechos de los niños y el derecho a la igualdad y a la no discriminación, entre otros.

Ejemplos de decisiones del Consejo asesor de contenido

El Consejo asesor de contenido [anuló](#) la decisión de Meta de no etiquetar un clip de audio probablemente manipulado de dos políticos kurdos iraquíes en el que hablaban sobre manipular las elecciones parlamentarias. El Consejo consideró que, para cumplir con las responsabilidades en materia de derechos humanos, Meta debería haber marcado el contenido de este caso, así como todas las apariciones del clip de audio, con la etiqueta "Alto riesgo".

La mayoría del Consejo asesor de contenido [confirmó](#) la decisión de Meta de dejar activas dos publicaciones que incluían videos en los que se confrontaba a una mujer transgénero por usar un baño de mujeres y a un atleta transgénero que ganó una carrera de atletismo.

Ejemplos de las medidas que tomamos a partir de las recomendaciones del Consejo asesor de contenido:

Siguiendo las [recomendaciones](#) en el caso de una [mujer iraní a la que se interpelló en la calle](#) y conforme a nuestras iniciativas para reducir la aplicación excesiva de las políticas, actualizamos nuestro enfoque para que dicha aplicación sea más reactiva. Esto ayudó a reducir los riesgos de tomar medidas contra expresiones figurativas o bromas que, potencialmente, no son amenazantes.

Siguiendo las [recomendaciones](#) de dar más detalles sobre sus decisiones respecto al contenido, Meta [aumentó](#) el nivel de detalle de las notificaciones que ya les enviaba a los usuarios. De este modo, proporcionó avisos más detallados cuando se restringe el contenido de un usuario en respuesta a un reporte oficial del Gobierno que indique que infringe las leyes locales.

Consulte los detalles en los [informes semestrales de Meta sobre el Consejo asesor de contenido](#) y nuestra [página sobre el impacto del Consejo asesor de contenido](#).

Impacto: aumentar la transparencia mediante notificaciones a los usuarios

+9%

T1 2024-T1 2025

Aumento en las tasas de notificación para la aplicación de las Normas comunitarias en Facebook

+16%

T1 2024-T1 2025

Aumento en las tasas de notificación para la aplicación de las Normas comunitarias en Instagram

~100%

en T1 2025

Tasa de notificaciones para contenido restringido por infringir las leyes locales en Instagram y Facebook



También ayudamos a respaldar el [liderazgo intelectual](#) del Consejo asesor de contenido en materia de libertad de expresión y evaluación de riesgos sistémicos, para que los reguladores pudieran aprender de la experiencia del Consejo. A lo largo del año, el Consejo participó en conversaciones sobre políticas en la Asamblea General de las Naciones Unidas y colaboró con la Comisión Africana de Derechos Humanos y de los Pueblos, entre otras partes interesadas.

*** Descargo de responsabilidad:** Existen numerosas diferencias entre los medios sociales y las películas. Al actualizar la configuración de contenido, no trabajamos con la MPA. Esta entidad no califica ningún contenido en Instagram, ni respalda ni aprueba de ninguna manera nuestra configuración de contenido. En cambio, nos basamos en las normas públicas de la MPA, las cuales ya resultan familiares para los padres, además de los comentarios de estos. Nuestros sistemas de moderación de contenido no funcionan igual que un consejo de clasificación de películas, por lo que es posible que la experiencia no sea exactamente la misma.



Anexo



Cómo se gestionan y regulan los derechos humanos en Meta

Nuestros expertos en derechos humanos guían la implementación de la [Política corporativa de derechos humanos](#), que está bajo la supervisión de nuestro director de asuntos globales y director jurídico.

Las tareas de los expertos en derechos humanos incluyen fomentar la integración de la política en programas, políticas 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 a fin de incorporar los derechos humanos a gran escala.

Capacitación en materia de derechos humanos para empleados de Meta

En Meta, cómo creamos es tan importante como qué creamos. Nuestra capacitación en derechos humanos destaca el impacto potencial y real de nuestros servicios, políticas y decisiones comerciales en cuanto a los 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.

Lanzamos nuestra capacitación *Bigger than Meta: Human Rights* en 2022, la cual continuó a lo largo de 2025. Nuestra capacitación en privacidad también respalda nuestros objetivos de formación en derechos humanos, ya que 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.

Enlaces a los informes mencionados

[Informes de prácticas comerciales responsables](#)

[Informes de sustentabilidad](#)

[Informe sobre derechos humanos de 2024](#), [Informe sobre derechos humanos de 2023](#), [Informe sobre derechos humanos de 2022](#), [Informe sobre derechos humanos de 2021](#)

[Declaración contra la esclavitud y la trata de personas de 2026](#)

[Informe sobre minerales en zonas de conflicto de 2025](#)

[Informes de transparencia de Meta](#)

[Informes regulatorios y otros informes de transparencia](#)

[Informes de amenazas adversas](#)

[Informes semestrales de Meta sobre el Consejo asesor de contenido](#)

