# Informe sobre derechos humanos

Meta

Información valiosa y medidas en 2024

# Índice


**Acerca de este informe** 03

**Resumen ejecutivo** 06

**Gestión de riesgos para los derechos humanos** 11

1. Libertad de opinión y expresión 12
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

**Velocidad de la innovación en materia de IA respetando los derechos humanos** 19

**Temas destacados** 25

2024: año electoral 25
- Preparación para las elecciones a gran escala 26
  - Gestionar los riesgos de influencia de la IA 26
  - Otras iniciativas de integridad de las elecciones 27
- Preparación para las elecciones de más alto riesgo 29
- Ejemplos de elecciones nacionales 29
  - Estados Unidos 29
  - México 30
  - India 31
  - Elecciones del Parlamento Europeo 32

Seguridad infantil y juvenil 33
- Protecciones integradas para adolescentes 33
- Lucha contra la sextorsión 36

Cómo nos preparamos para afrontar una crisis y respondemos ante estas situaciones 37
- Sudán 39
- Oriente Medio 41
- Bangladesh 42
- Georgia 43

Ciberseguridad 44

**Participación de partes interesadas** 46

Foros de la comunidad 51

Socios de confianza 52
- Caso de éxito: Informes elaborados a partir de estadísticas de Siria 54
- Caso de éxito: Mitigación de riesgos para los agentes civiles en Venezuela 55
- Caso de éxito: Los socios de confianza abordan las acusaciones de blasfemia y el lenguaje hostil en Pakistán 56

Organizaciones internacionales 58

**Transparencia y resarcimiento** 60

**Anexo** 64

Cómo Meta gestiona y rige el trabajo relativo a los derechos humanos 65

Capacitación en materia de derechos humanos para empleados de Meta 65

Enlaces a los informes mencionados 65


# 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 2024. Incluimos información sobre los servicios y productos de Meta, como Facebook, Messenger, Instagram, WhatsApp, Threads y Reality Labs.

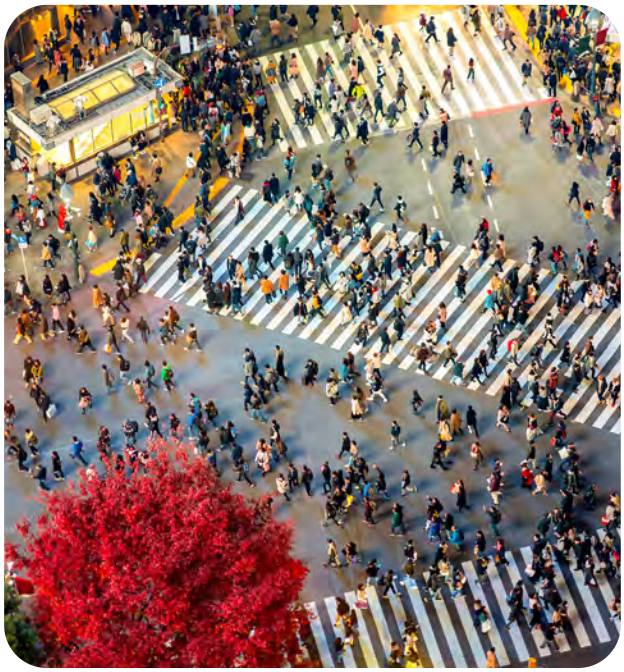
Se basa en el trabajo realizado por Meta en pro de los derechos humanos y refleja el progreso conseguido respecto de nuestros compromisos con los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas y nuestra Política corporativa de derechos humanos. En el informe, se explica cómo aplicamos estos principios en la empresa en 2024 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 impactos[1] 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 medidas de mitigación que emprendimos en 2024.[2]

En 2024, los derechos humanos seguían siendo un tema de fundamental importancia tanto para nuestra empresa como para nuestras partes interesadas. Nos esforzamos por proporcionar una muestra representativa de nuestra labor en diversos equipos y en colaboración con partes interesadas de todo el mundo.

## Políticas y progreso realizado

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 basada en indicadores
- Informe de prácticas comerciales responsables
- Centro de transparencia
- Informe de sostenibilidad
- Informe sobre el cambio climático del CDP
- Pacto Mundial de las Naciones Unidas

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. 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.

[2] No incluye los cambios a la política de contenido y las demás modificaciones que anunciamos en enero de 2025, cuando actualizamos nuestra Política de conducta que incita al odio, antes conocida como Política de lenguaje que incita al odio, para abordar inquietudes respecto de la aplicación excesiva de políticas y permitir una mayor libertad de expresión.

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.[3]

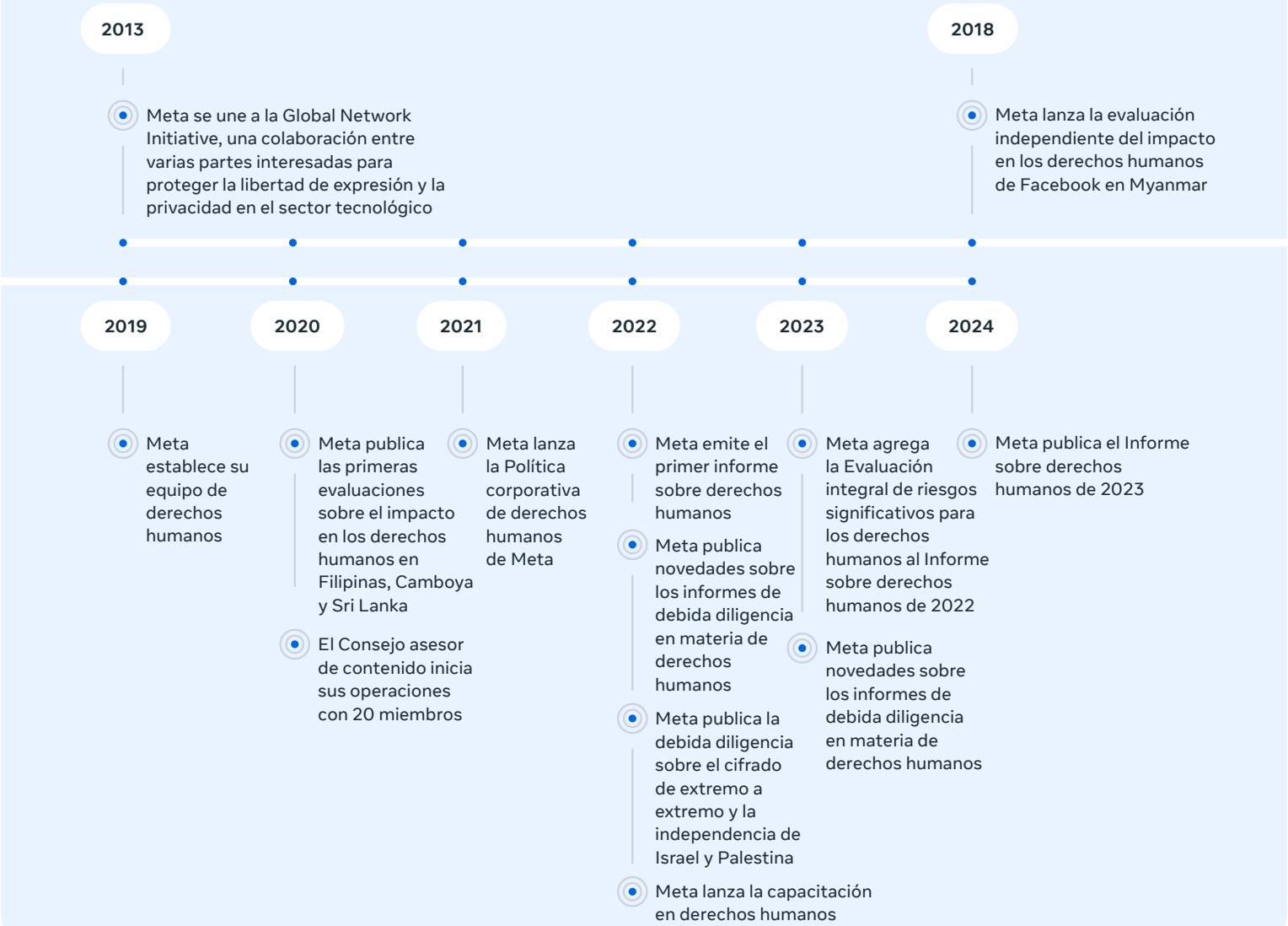
[3] El debate sobre moderación de contenido y las medidas relacionadas en Facebook e Instagram que se detallan en este informe no se aplican a WhatsApp. Asimismo, a menos que se indique expresamente que una política o medida se aplica a WhatsApp, se considera que no se le aplica. Además, si bien muchas 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 cuatro informe anual sobre derechos humanos de Meta. En él se brinda información valiosa sobre la labor que realizó Meta en 2024 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 (UNGP).

# Cronología de los derechos humanos

En la siguiente tabla, se detalla nuestro recorrido en materia de derechos humanos y cómo evolucionó nuestro trabajo desde que el Consejo de Derechos Humanos de la ONU adoptó los UNGP en 2011.




## Evaluación de riesgos significativos

Leer más

Nuestras prioridades en 2024 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.

## Acelerar la innovación en materia de IA

Los avances en el ámbito de la inteligencia artificial (IA) se aceleraron en 2024. Nuestra visión es crear una superinteligencia personal ampliamente disponible, de modo que todos puedan sacar partido de ella.

Se extendió el uso de apps de IA generativa, que fueron transformando cada vez más la manera en que nos comunicamos, aprendemos, creamos y trabajamos. Seguimos fomentando un enfoque abierto respecto de la IA que pueda mejorar los derechos humanos. Este enfoque contribuyó a que las personas accedan a información y gocen de libertad de expresión, así como a fomentar el derecho a la igualdad y la no discriminación, incluso mediante una mejora de la accesibilidad y una mayor inclusión en el lenguaje.

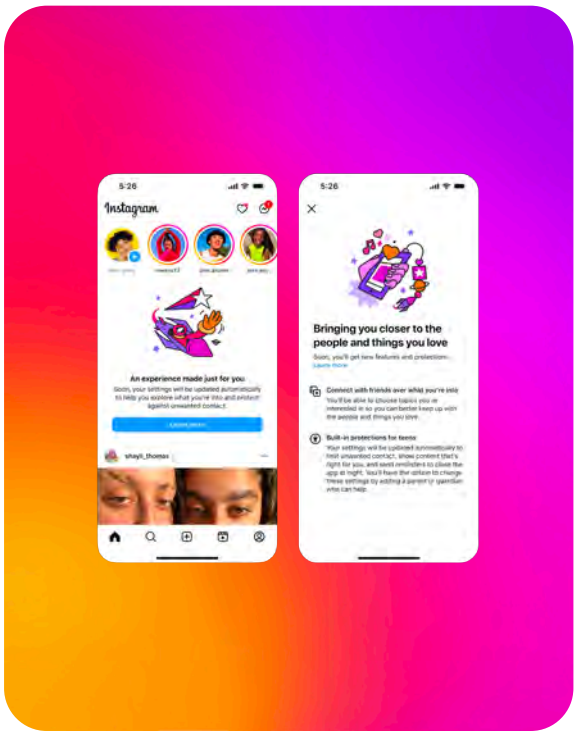
Leer más

## 2024: año electoral

En 2024, fue el año en que se realizaron más elecciones en la historia. Se celebraron elecciones nacionales en más de 70 países, donde vive más de la mitad de la población mundial. Aproximadamente, 2.000 millones de personas tuvieron la posibilidad de votar. Nos centramos en fomentar los derechos a la libertad de expresión, la participación en procesos políticos y el acceso a la información en aquellos países donde se realizaron elecciones.

Nuestro enfoque maduró en el transcurso de cientos de elecciones que se llevaron a cabo en los últimos años. Implicó iniciativas para gestionar los riesgos que supone la IA, aplicar nuestras políticas en situaciones de interferencia electoral o en el censo, interrumpir redes adversas, aumentar la transparencia de la publicidad política y conectar a los votantes con información confiable. En este informe, se incluyen ejemplos de los Estados Unidos, México, la India y la Unión Europea.

Leer más




## Seguridad infantil y juvenil

Continuamos con nuestro compromiso con la seguridad infantil y juvenil. Entre otras iniciativas, nuestra labor en 2024 incluyó el lanzamiento de las cuentas de adolescente de Instagram, una nueva experiencia para adolescentes con los padres como guía. Las cuentas de adolescente tienen protecciones incorporadas que imponen límites respecto de quién puede contactar a los adolescentes y del contenido que estos ven. Además, las cuentas de este tipo ofrecen vías para gestionar el tiempo que los adolescentes pasan en la app y les proporcionan nuevas formas de explorar sus intereses. Nuestras iniciativas buscaban un equilibrio entre respaldar y fomentar la autonomía de los jóvenes y reconocer los derechos y las obligaciones de padres y tutores. Las desarrollamos en consonancia con las recomendaciones de expertos y el principio de las capacidades evolutivas del niño de la Convención sobre los Derechos del Niño de la ONU.

Leer más

## Respuesta a crisis

Seguimos integrando los principios de derechos humanos en cuanto a cómo nos preparamos para afrontar una crisis y respondemos a ella. Nuestro Protocolo de la política de crisis guía nuestro uso inmediato de tácticas para mitigar posibles daños. En 2024, designamos 19 situaciones en todo el mundo en virtud de este protocolo. En este informe, proporcionamos ejemplos de nuestra respuesta ante crisis en Bangladesh, Georgia, Oriente Medio y Sudán.

Leer más

## Participación de partes interesadas

Nuestra Política corporativa de derechos humanos respalda nuestra colaboración proactiva con partes interesadas. En 2024, nos conectamos con una diversidad de partes interesadas para transmitir el enfoque de la empresa respecto de los problemas relacionados con la expresión, el contenido de odio, la información errónea y la privacidad. Estas partes interesadas incluían a diversos grupos de derechos humanos, comunidades vulnerables, miembros de la sociedad civil, académicos, grupos de expertos y autoridades reguladoras. Los temas clave incluían nuestro enfoque respecto de una IA responsable y la integridad de las elecciones, así como nuestras señales de designación de personas y organizaciones peligrosas, y eventos violentos.

En 2024, realizamos seis foros de políticas, donde expertos en la materia de Meta comparten diversos puntos de vista y debaten posibles cambios en las Normas comunitarias y las Normas de publicidad. Asimismo, realizamos foros de la comunidad para aprovechar los aportes del público respecto de problemas donde había dilemas contrapuestos y ninguna respuesta clara. Estos nos permitieron mejorar los productos y anticiparnos a los riesgos potenciales que suponen las tecnologías emergentes, así como brindar mayor participación en la toma de decisiones a personas y entidades externas a la empresa.

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. Asimismo, exploramos cómo fortalecer los canales de escalamiento relevantes. Sus conocimientos fueron de particular valor durante el intenso ciclo electoral de 2024 y en situaciones de agudo malestar. También proporcionaron información valiosa e identificaron contenido potencialmente infractor en Bangladesh, Brasil, Costa de Marfil, Francia, Grecia, India, Indonesia, Kenia, México, Nigeria, Pakistán, Región de Kurdistán, República Democrática del Congo, Senegal, Siria, Sudáfrica y Venezuela, entre otros países y regiones.

Leer más

## Consejo asesor de contenido

En 2024, el Consejo asesor de contenido consideró casos relativos a nuestras iniciativas para respetar los derechos humanos, incluido el derecho a la libertad de expresión, a la salud y a la no discriminación, entre otros temas. El Consejo asesor de contenido es un organismo independiente que 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. Este Consejo proporciona reglas vinculantes respecto de si conservar o eliminar el contenido en cuestión. En respuesta a una de sus recomendaciones, Meta evaluó la puntualidad y efectividad de las respuestas ante contenido reportado mediante el Programa de socios de confianza.

Leer más

## Administración de solicitudes gubernamentales

A lo largo del año, 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. En 2024, publicamos casos de éxito relacionados con el discurso político en Alemania, Brasil, India, Irak, Israel, Singapur y Turquía.

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.

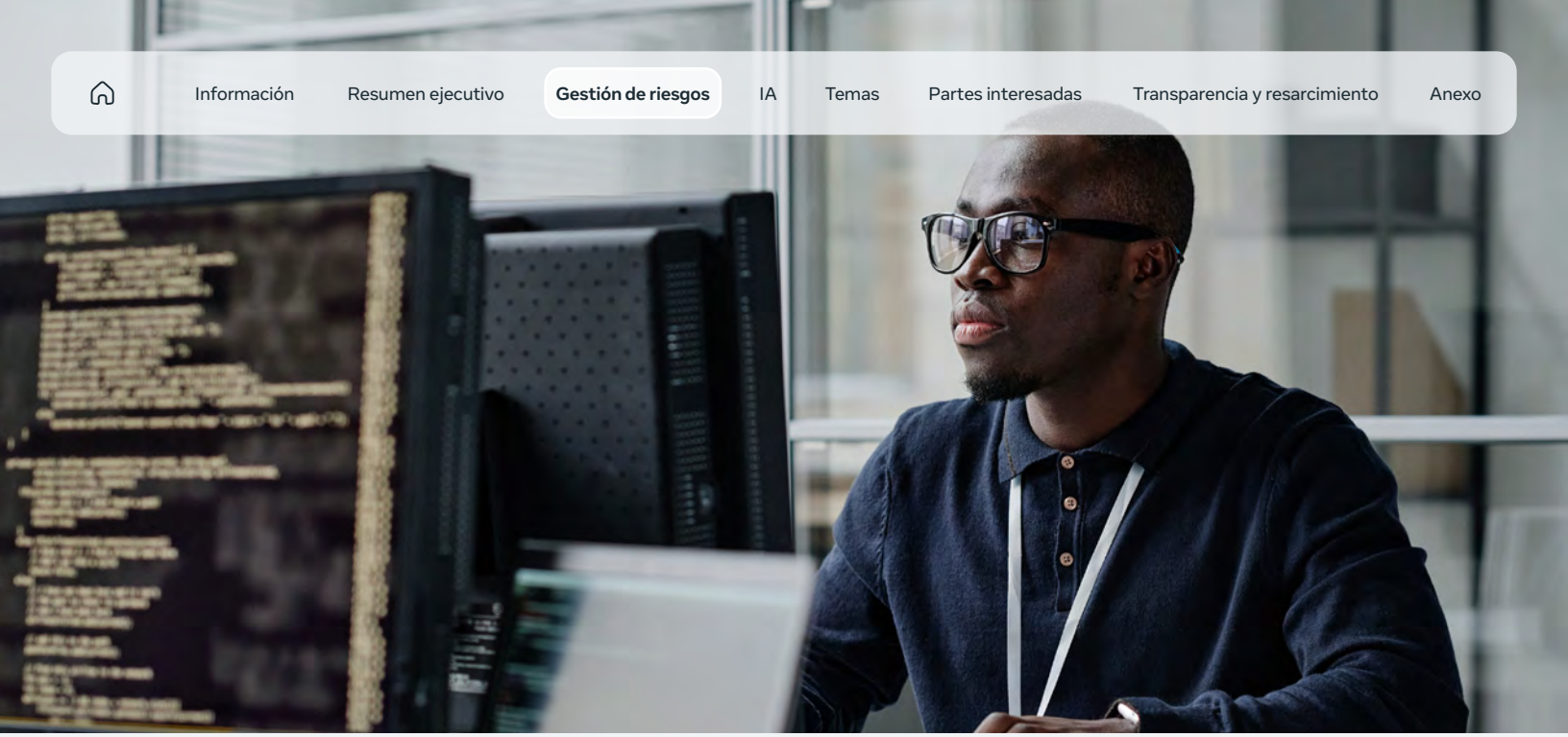
Una vez implementados los procesos para abordar estos riesgos, siempre queda cierto nivel de riesgo, que se conoce como riesgo "residual". Si bien hay riesgos residuales en todos los sistemas de gestión de riesgos, persisten aquellos asociados con las tecnologías digitales y su impacto en los derechos humanos, debido a la naturaleza dinámica y de rápida evolución de dichas tecnologías y al alto nivel de actividad externa.

En la tabla que figura en las páginas siguientes, se detallan los riesgos significativos para los derechos humanos que detectamos conforme se definen en nuestra Evaluación integral de riesgos significativos para los derechos humanos (CSRA) de 2022, que publicamos en nuestro informe sobre derechos humanos de 2022. Esta tabla brinda ejemplos ilustrativos de cómo abordamos los riesgos potenciales en 2024. 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 solicitar, 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 2024 |
|---|---|
| Las políticas de moderación de contenido de Meta y su aplicación pueden limitar la libertad de expresión. | Seguimos desarrollando nuestras políticas tomando la libertad de expresión como brújula que guía nuestro camino. En 2024, realizamos varios foros de políticas cuyo objetivo era desarrollar una valoración matizada de los desafíos en cuanto a la libertad de expresión en diversas áreas. |
| El gobierno impone límites excesivos sobre el contenido. | Luchamos por cumplir los compromisos que asumimos con la Global Network Initiative (GNI), que incluyen informar sobre nuestras respuestas a las solicitudes gubernamentales de datos o restricciones de contenido (aquí y aquí). El enfoque que empleamos para responder a estas solicitudes se detalla en nuestro Informe sobre derechos humanos de 2023. 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, presentar una apelación o no tomar ninguna medida. En 2024, entre los casos de éxito relacionados con la transparencia en cuanto al discurso político para destacar, se encontraban aquellos en Alemania, Brasil, India, Irak, Israel, Singapur y Turquía. |
| Las interrupciones del servicio de internet y los bloqueos en medios sociales evitan que las personas ejerzan su derecho a la libertad de expresión e impiden que intercambien noticias e información vitales. | Para evitar que se bloqueen medios sociales y mensajes, podemos atender solicitudes gubernamentales legales, sin perder de vista nuestros compromisos con la GNI de respetar la libertad de expresión. Asimismo, continuaremos brindando la función de WhatsApp by Proxy para personas que no pueden conectarse a nuestras apps directamente. |



# 2. Privacidad

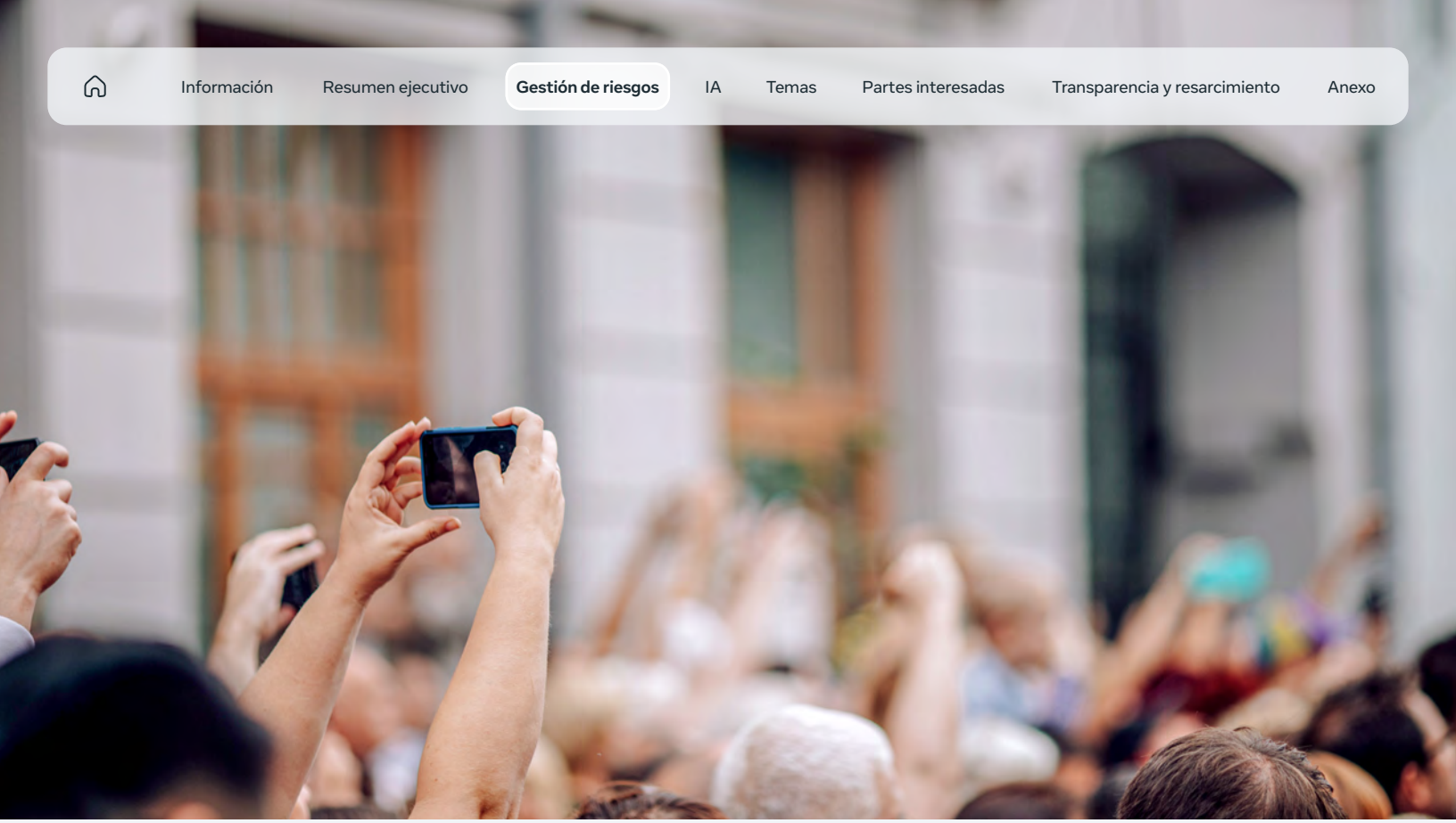
El derecho a la privacidad es condición necesaria para que se cumplan los demás derechos humanos, como la libertad de expresión, la libertad de reunión y asociación, y la libertad de culto y religión. Uno de los principios fundamentales que se detallan en nuestra Política corporativa de derechos humanos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

| Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| Los modelos de IA generativa pueden implicar el tratamiento de datos personales de modos que las personas no prevén ni comprenden. | Somos transparentes sobre cómo Meta usa la información en modelos y funciones de IA generativa, y contamos con un proceso de revisión de privacidad interno para un uso responsable de los datos, incluida la IA generativa. Hay disponible una actualización de nuestro progreso en materia de privacidad en 2024 aquí y aquí.<br><br>Leer más |
| El contenido o los comportamientos en las apps de Meta pueden afectar de manera adversa los derechos de privacidad y protección de datos. | En octubre de 2024, Meta volvió a introducir la tecnología de reconocimiento facial en Facebook e Instagram para que las personas recuperaran cuentas comprometidas y evitaran los fraudes relacionados con el apoyo falso de celebridades. Para equilibrar los posibles riesgos de privacidad con los de integridad, brindamos a las figuras públicas cuya imagen se usa de manera indebida para engañar a otras personas la opción de formar parte o no participar del programa. |

# 3. Igualdad y no discriminación

El derecho a la igualdad y la no discriminación brinda igualdad de protección contra cualquier tipo de 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 | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| La moderación en algunos idiomas y dialectos puede suponer un mayor desafío. | Diseñamos e implementamos nuevos mecanismos para remitir contenido en idioma árabe por dialecto, a fin de que la moderación resulte más eficiente y precisa, incluido en Sudán. El nuevo sistema detecta el contenido y prioriza derivarlo a moderadores que es más probable que comprendan ese dialecto árabe en particular. |
| Hay contenido que afecta de manera adversa a la igualdad y la no discriminación (p. ej., conducta que incita al odio). | En función de investigaciones, colaboraciones externas e indagaciones en nuestras plataformas, actualizamos nuestra Política de conducta que incita al odio en cuanto al contenido que ataca a los "sionistas".<br><br>Durante el entrenamiento de nuestros modelos de IA, probamos datos de entrenamiento para contenido o propiedades que podrían aumentar el riesgo de que se genere contenido potencialmente dañino, por ejemplo, si un conjunto de datos representa a diversos grupos demográficos. |



# 4. Vida, libertad y seguridad de la persona

El derecho a la vida, la libertad y la seguridad de la persona concierne no estar sujeto a 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 | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| Personas malintencionadas que:<br>• Explotan las apps y los servicios de Meta para coordinar daños online y offline.<br>• Usan indebidamente apps y servicios para realizar ciberataques o phishing.<br>• Amenazan y acosan a defensores de los derechos humanos, activistas y otros grupos vulnerables. | La Política de organización de actos dañinos y promoción de la delincuencia de Meta prohíbe facilitar, organizar, promover o aceptar ciertas actividades criminales o dañinas. En 2024, proporcionamos pautas respecto de los prisioneros de guerra en la política, de modo que los revisores de contenido pudieran eliminar contenido infractor con mayor eficacia a gran escala, incluido en Sudán.<br><br>Seguimos respaldando el Fondo para defensores de derechos humanos y rediseñamos nuestro Programa de socios de confianza para 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 | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| Los niños pueden quedar expuestos a contenido no deseado e inapropiado o a comportamientos depredadores. | Lanzamos las cuentas de adolescente de Instagram, una nueva experiencia para los adolescentes con protecciones incorporadas, cuya guía son los padres.<br><br>Leer más |

# 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 | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| El contenido 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:<br>• Agentes malintencionados coordinados que interfieren en procesos electorales<br>• Amenazas de daño y violencia en la vida real contra candidatos<br>• Iniciativas individuales para disuadir a las personas de votar, aumento de la cantidad de spam, injerencia extranjera o reportes de contenido que infringe nuestras políticas | Las elecciones eran una prioridad en 2024. Nos preparamos para las elecciones a gran escala, incluido para las de más alto riesgo, y ayudamos a los votantes a encontrar información, entre otras iniciativas.<br><br>Leer más |

# 7. Libertad de reunión y asociación

El derecho a la libertad de reunión y asociación es fundamental para la democracia y es mutuamente dependiente 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 2024 |
|---|---|
| Debido al contenido o al comportamiento no auténtico coordinado en las plataformas de Meta, algunas personas pueden sentirse incapaces de reunirse con libertad en las apps de Meta u offline. | Implementamos nuestro Protocolo de la política de crisis a fin de dotar de recursos nuestras iniciativas para abordar el contenido infractor relacionado con manifestaciones masivas, por ejemplo, en Bangladesh y Georgia.<br><br>Asimismo, nos preparamos mucho tiempo antes de las elecciones para reducir el riesgo de que se difunda contenido infractor que podría hacer sentir a las personas en peligro durante las elecciones y una vez finalizadas.<br><br>Threads se sumó al fediverso, una red mundial de servidores de medios sociales, lo que permitió a las personas ampliar sus comunidades y llegar a públicos nuevos. |

# 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 confiable, permite que personas con problemas de salud similares se conecten entre sí y les brinda herramientas para tomar decisiones fundamentadas sobre su salud y bienestar.

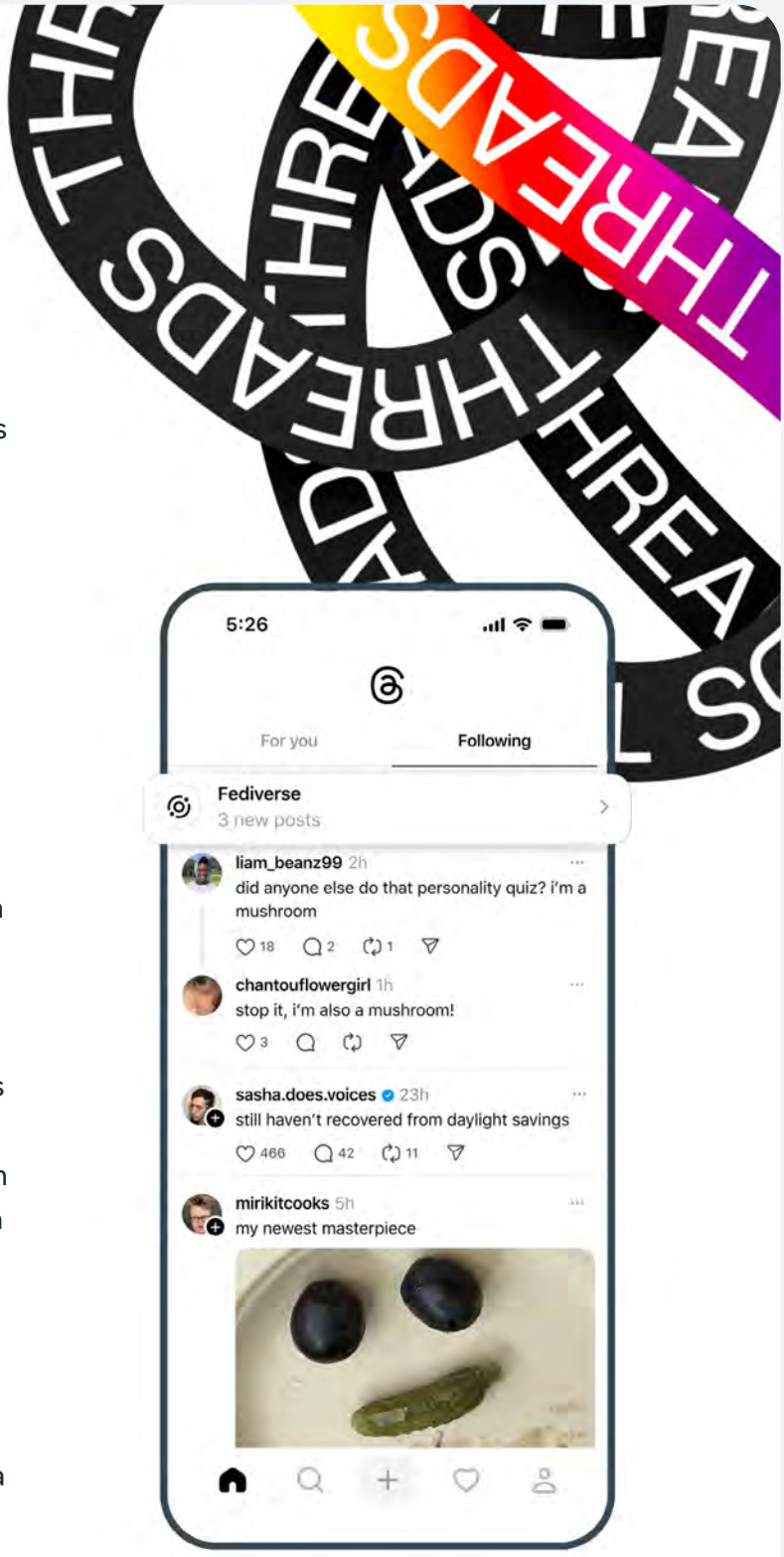
| Ejemplos de posibles riesgos significativos inherentes para los derechos humanos identificados en la CSRA | Ejemplos de cómo Meta abordó los riesgos potenciales en 2024 |
|---|---|
| Contenido que infringe las normas que incita a cometer daños en la vida real o se diseñó con ese fin. | En conjunto con Snap y TikTok, lanzamos Thrive, un programa de uso compartido de señales intersectorial diseñado para evitar la difusión de contenido relacionado con suicidio y autoagresión.<br><br>Realizamos un foro de políticas sobre contenido comercial en el que se tuvieron en cuenta riesgos de salud y seguridad reconocidos por la normativa vigente.<br><br>Actualizamos nuestras Normas comunitarias y las Normas de publicidad para que hagan referencia a los productos retirados del mercado. |

## Nuevos productos y servicios

Como mencionamos en nuestro Informe sobre derechos humanos de 2023, Meta se esfuerza por respetar los derechos humanos en el diseño y desarrollo de nuestros productos y servicios.

En 2024, Threads se sumó al fediverso, una red mundial de servidores de medios sociales. Si un usuario decide comenzar a compartir contenido en el fediverso, personas de diferentes plataformas (como Mastodon o Flipboard) pueden seguir el contenido de Threads de ese usuario e interactuar con él, incluso aunque no tengan un perfil de Threads. De esta manera, las personas pueden ejercer su derecho a la libertad de expresión, reunión y asociación, ya que llegan a nuevos públicos, amplían sus comunidades y se suman al debate público sobre temas que les interesan. Esto también permite crear un ecosistema de información más diverso.

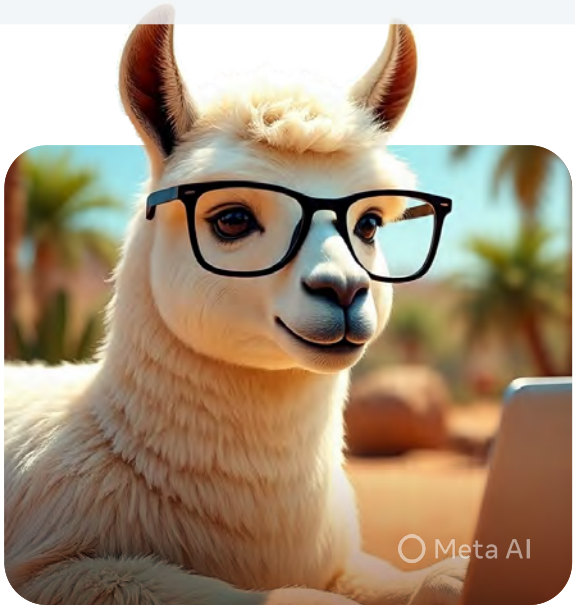
También brindamos a las personas que usan Threads información educativa mediante una sección especializada en nuestro servicio de ayuda y les ofrecemos una nueva guía del fediverso en nuestro centro de privacidad sobre cómo la descentralización y la interoperabilidad afectan a la privacidad.



# Velocidad de la innovación en materia de IA respetando los derechos humanos

Los avances en el ámbito de la inteligencia artificial (IA) se aceleraron en 2024. Se extendió el uso de apps y herramientas de IA generativa, que fueron transformando cada vez más la manera en que nos comunicamos, aprendemos, creamos y trabajamos.
En Meta, reconocemos que el veloz desarrollo y la rápida adopción de la IA conlleva ventajas y riesgos importantes, y a menudo desconocidos, para los derechos humanos.

Nuestra visión a largo plazo es crear una superinteligencia personal ampliamente disponible, de modo que todos puedan sacar partido de ella.




En 2024, lanzamos nuestros macromodelos lingüísticos abiertos Llama 3, Llama 3.1, Llama 3.2 y Llama 3.3. Lanzamos también nuestro asistente de Meta AI, que integramos en nuestras tecnologías. Meta AI Studio debutó como una plataforma para crear personajes de IA personalizados, un paquete de herramientas de IA generativa ayudó a los anunciantes a desarrollar su negocio y Meta AI se integró a nuestros lentes Ray-Ban Meta. Seguimos realizando y publicando investigaciones sobre IA de vanguardia, incluidos nuestros modelos Movie Gen que generan videos y ofrecen funciones de edición precisas basadas en instrucciones y nuestro modelo Video Seal para insertar marcas de agua permanentes en videos generados con IA, entre otros avances.

Para finales de 2024, los desarrolladores habían descargado nuestros modelos abiertos Llama más de 650 millones de veces y Meta AI contaba con casi 600 millones de usuarios activos por mes en todo el mundo. Estas cifras demuestran que son los más adoptados del mundo. Esta enorme base de usuarios conformada por desarrolladores y usuarios finales pone de manifiesto nuestra responsabilidad de desarrollar la IA de un modo que respete los derechos humanos.

## Nuestro enfoque abierto

Creemos que la IA de open source es fundamental para garantizar que todos disfruten de los beneficios que suponen los avances en el ámbito de la IA. Como mencionamos en nuestro Informe sobre derechos humanos de 2023, un enfoque abierto supone importantes beneficios para los derechos humanos. Los modelos de IA de open source:

- De manera inherente, resisten más a la censura y a otras restricciones impuestas al derecho a la libertad de expresión porque se pueden descargar y ejecutar offline. Esto reduce el impacto de las posibles exigencias gubernamentales que reclaman restringir los resultados que arrojan luego de su lanzamiento.
- Favorecen la adaptación y el perfeccionamiento para reflejar el contexto y los matices locales, de conformidad con el derecho a la igualdad, lo que amplía la accesibilidad y la inclusión en el lenguaje.
- Permiten que los desarrolladores creen con mayor facilidad modelos más pequeños y eficientes, que pueden beneficiar a comunidades tradicionalmente desfavorecidas, y respaldar así los derechos económicos, sociales y culturales.
- Respaldan investigaciones críticas sobre medidas de seguridad y protección en el ámbito de la IA permitiendo que todos supervisen los modelos en busca de posibles riesgos, gracias a lo cual es posible mitigar el potencial impacto negativo en los derechos humanos.

Nuestro enfoque abierto ya nos muestra beneficios tangibles. Nuestro programa Llama Impact Grants, que lanzamos en 2023, continuó en 2024. Este programa, junto con Llama Impact Innovation Awards en 2024, respalda y destaca los casos de uso con impacto social positivo de nuestros modelos abiertos.

Por ejemplo, los desarrolladores usaron Llama para crear el modelo Vax-Llama, un servicio de chatbot diseñado para brindar información precisa sobre vacunas, con el objeto de que lo adopten proveedores de atención médica de todo el mundo. También se utilizó para el proyecto Llama-Suho, una iniciativa con la que perfeccionamos Llama con datos específicos de Corea para mejorar las medidas de protección de la IA en el contexto de este país.

# Protecciones focalizadas

Mantenemos el compromiso de desarrollar e implementar productos de IA de vanguardia, sin perder de vista las normas de derechos humanos y las medidas de protección contra el uso indebido.

Nuestra Política corporativa de derechos humanos menciona explícitamente la capacidad de aplicación de nuestros compromisos con la IA y los derechos humanos.

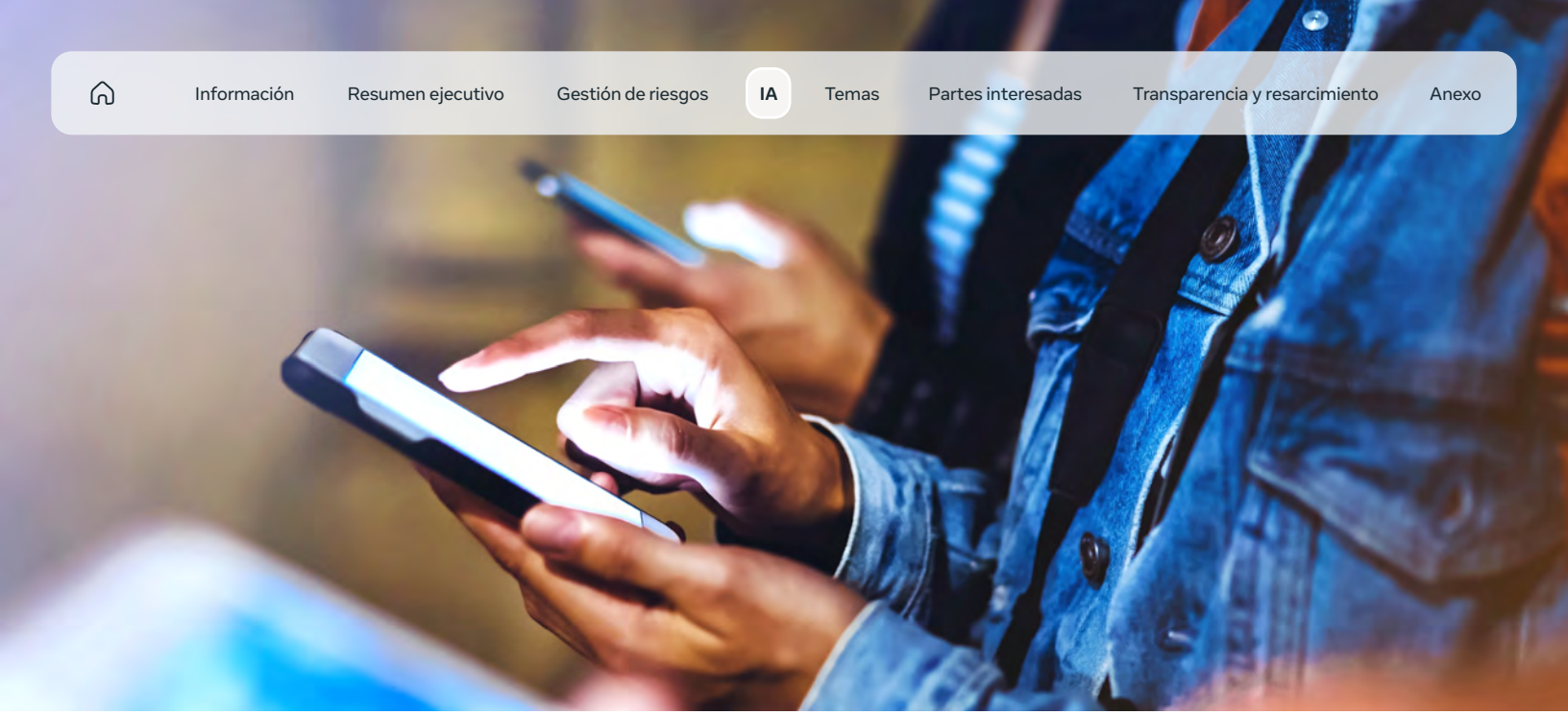
Con el lanzamiento de Llama 3 en abril de 2024, comenzamos a hacer hincapié en un enfoque basado en sistemas respecto de las medidas de protección de la IA. Un sistema de esta naturaleza brinda a los desarrolladores más flexibilidad para aplicar los niveles adecuados de protección en diferentes casos de uso y públicos. Por ejemplo, brindamos protecciones para determinados tipos de discursos potencialmente ofensivos, pero legales, como medidas de mitigación opcionales en el sistema. Esto se suma a la continua incorporación de medidas de protección de referencia contra la generación de contenido de explotación infantil en nuestros modelos fundacionales.

Consideramos que este enfoque basado en sistemas favorece un equilibrio adecuado entre libertad de expresión y otros derechos humanos.

## Sistemas de seguridad de IA


Como parte de nuestro enfoque basado en sistemas, ofrecemos tres herramientas clave como open source (Llama Guard, Prompt Guard y Code Shield) que los desarrolladores pueden personalizar y usar en conjunto o de forma independiente para implementar medidas de protección contra el uso indebido.

Nuestra Guía de uso para desarrolladores brinda pautas detalladas sobre cómo implementar de manera responsable nuestros modelos y sistemas de seguridad fundacionales en una variedad de contextos. Seguimos imponiendo nuestra Política de uso aceptable a las implementaciones de nuestros modelos abiertos.



Además de proporcionar estas herramientas en 2024, también dimos importantes pasos para mitigar los riesgos asociados con nuestras propias implementaciones de la IA generativa. Entre las prácticas que adoptamos en 2024, se encuentran las siguientes:

- Prácticas exhaustivas de red-teaming en nuestros modelos y productos propios antes de su lanzamiento para identificar y mitigar posibles riesgos, incluidos aquellos relacionados con el posible impacto negativo en los derechos humanos.
- Actualización de nuestras técnicas de gestión de contenido multimedia manipulado en función de comentarios del Consejo asesor de contenido independiente, que incluyen agregar la etiqueta "Información de IA" y contexto a una gama más amplia de contenido de video, audio e imágenes, y exigir que los creadores indiquen expresamente que están usando IA.
- Perfeccionamiento de los procesos y las pautas internas que usamos para probar resultados del modelo aceptables, a fin de reflejar mejor casos de uso del mundo real y actuar en consonancia con las normas de derechos humanos internacionales.

Reconocemos también que las medidas de protección de la IA requieren de una colaboración intersectorial y de diversas partes interesadas. En febrero de 2024, junto con otras empresas afines del sector, firmamos el Pacto Tecnológico contra el Uso Engañoso de la IA en las Elecciones, en el que nos comprometimos a trabajar para prevenir que contenido generado con IA engañoso interfiera en los procesos electorales mundiales. En mayo de 2024, nos unimos a Frontier Model Forum, un organismo que cuenta con el respaldo del sector y se dedica a lograr progresos relativos a la seguridad de los modelos de IA avanzados.

## Abordaje de falsos rechazos

Un falso rechazo ocurre cuando un modelo se niega a producir el resultado solicitado en respuesta a una indicación válida, con frecuencia debido a las protecciones de seguridad bien intencionadas del modelo. Por ejemplo, un modelo puede negarse por error a debatir sobre una pieza de la literatura clásica que contiene un estereotipo ofensivo o insultos, o a responder una pregunta básica sobre química de nivel de escuela secundaria a causa de las medidas de protección diseñadas para evitar brindar ayuda con la creación de explosivos químicos, biológicos, radiológicos, nucleares y de alta potencia. Si bien la seguridad del modelo es importante y los rechazos tal vez sean necesarios para limitar la generación de contenido dañino, los falsos rechazos pueden tener un impacto negativo en los derechos a la libertad de expresión, el acceso a la información, entre otros.

Con Llama 3 como punto de partida, realizamos una importante labor para reducir los falsos rechazos de Llama y Meta AI, y logramos un progreso significativo en el transcurso de 2024.

## Internacionalización responsable

En 2024, lanzamos Meta AI en más de otros 40 países y en varios idiomas nuevos, incluidos alemán, árabe, español, indonesio, filipino, francés, hindi, italiano, portugués, tailandés y vietnamita.

Antes del lanzamiento en cada país e idioma, evaluamos los posibles riesgos para los derechos humanos y aplicamos prácticas de red-teaming específicas del contexto, medida habitual para mitigar comportamientos poco seguros en macromodelos lingüísticos.

No todos los países en los que está disponible Meta AI cuentan con una legislación nacional que incluya medidas de protección sólidas para salvaguardar la libertad de expresión. Como parte de nuestra labor en materia de internacionalización en 2024, desarrollamos un enfoque basado en derechos humanos para responder a las solicitudes gubernamentales que exigen restringir o limitar el resultado que proporciona Meta AI. Para ello, nos basamos en nuestras políticas de larga trayectoria y actuamos de conformidad con nuestros compromisos como miembro de Global Network Initiative y nuestra Política corporativa de derechos humanos.

## Colaboración con partes interesadas

Todo espacio en el que la tecnología avanza tan rápido como la IA supone desafíos desconocidos para la colaboración con partes interesadas. A lo largo de 2024, nos centramos en brindarles información y solicitar sus comentarios significativos.

Entre nuestras iniciativas figuran las siguientes:

- Llevamos a cabo mesas redondas sobre IA en los Estados Unidos para recopilar comentarios de grupos interdisciplinarios acerca del lanzamiento de productos y modelos, incluida la opinión de expertos de Estados Unidos, Brasil, Bruselas, Jordania, México y diversas partes de África.
- Compartimos los hallazgos obtenidos de los foros de la comunidad realizados en los Estados Unidos, Brasil, Alemania y España para explorar los principios de los chatbots con IA generativa, con la colaboración de Deliberative Democracy Lab de la Universidad de Stanford.
- Realizamos una serie de talleres Open Loop diseñados para abordar las complejidades y aprovechar las oportunidades de la IA de open source. En estos talleres se reunieron legisladores, líderes del sector, académicos y representantes de la sociedad civil de todo el mundo para forjar, en un enfoque colaborativo, políticas de IA responsable y efectiva.
- Junto con el Foro de las Naciones Unidas sobre las Empresas y los Derechos Humanos n.º 13 celebrado en Ginebra, desarrollamos y dirigimos una simulación interactiva con diversas partes interesadas sobre la debida diligencia en materia de derechos humanos para productos de IA generativa, en la que compartimos nuestro enfoque y logramos un mayor entendimiento mutuo respecto de los riesgos y desafíos que plantean.

A medida que seguimos innovando en el campo de la IA, sostenemos nuestro compromiso de lograr que las partes interesadas participen y brinden asesoramiento de un modo inclusivo y sólido en todo el mundo.

Leer más

# Temas destacados

## 2024: año electoral

En 2024, fue el año en que se realizaron más elecciones en la historia. Se celebraron elecciones nacionales en más de 70 países, donde vive más de la mitad de la población mundial, y aproximadamente 2.000 millones de personas tuvieron la posibilidad de votar.

Meta reconoce lo importante que es dar lugar a que las personas ejerzan sus derechos a la libertad de expresión, al voto y a participar en los asuntos públicos. Durante el año, nos centramos principalmente en nuestro trabajo en el ámbito electoral. Nos preparamos para el alcance, la difusión y la frecuencia de las elecciones y trabajamos para mitigar riesgos relacionados para los usuarios, incluidos aquellos que entrañan aumentar el uso de la IA.

En las siguientes páginas, revisaremos las iniciativas que emprendimos en 2024 y brindaremos resúmenes ilustrativos de las elecciones en la Unión Europea, la India, México y los Estados Unidos.

## Preparación para las elecciones a gran escala

Meta realizó cambios a nuestro enfoque central respecto de las elecciones en el transcurso de los últimos años. Los implementamos en todos los países donde se usan nuestros servicios, y adaptamos nuestra estrategia conforme a las necesidades y los riesgos locales. Como parte de nuestros preparativos para las elecciones de 2024, un equipo especializado se encargó de las iniciativas interempresariales, que incluían a expertos de nuestros equipos de inteligencia, ciencia de datos, productos e ingeniería, investigación, operaciones, contenido, derechos humanos, política pública y legal. A lo largo del año, aspiramos a brindar herramientas que permitieran a las personas expresar su opinión, votar y resultar elegidas para ocupar un cargo.

Nuestro enfoque incluyó iniciativas para gestionar los riesgos que supone la IA, aplicar nuestras políticas en situaciones de interferencia electoral, interrumpir redes adversas, brindar transparencia en la publicidad política y conectar a los votantes con información confiable. Asimismo, evaluamos la cobertura lingüística adecuada de los clasificadores y revisores humanos en los países que celebraban elecciones con el objetivo de tomar medidas sobre contenido que infringe las políticas conforme a nuestras iniciativas en este respecto. Algunos aspectos destacados de nuestra labor incluyeron:

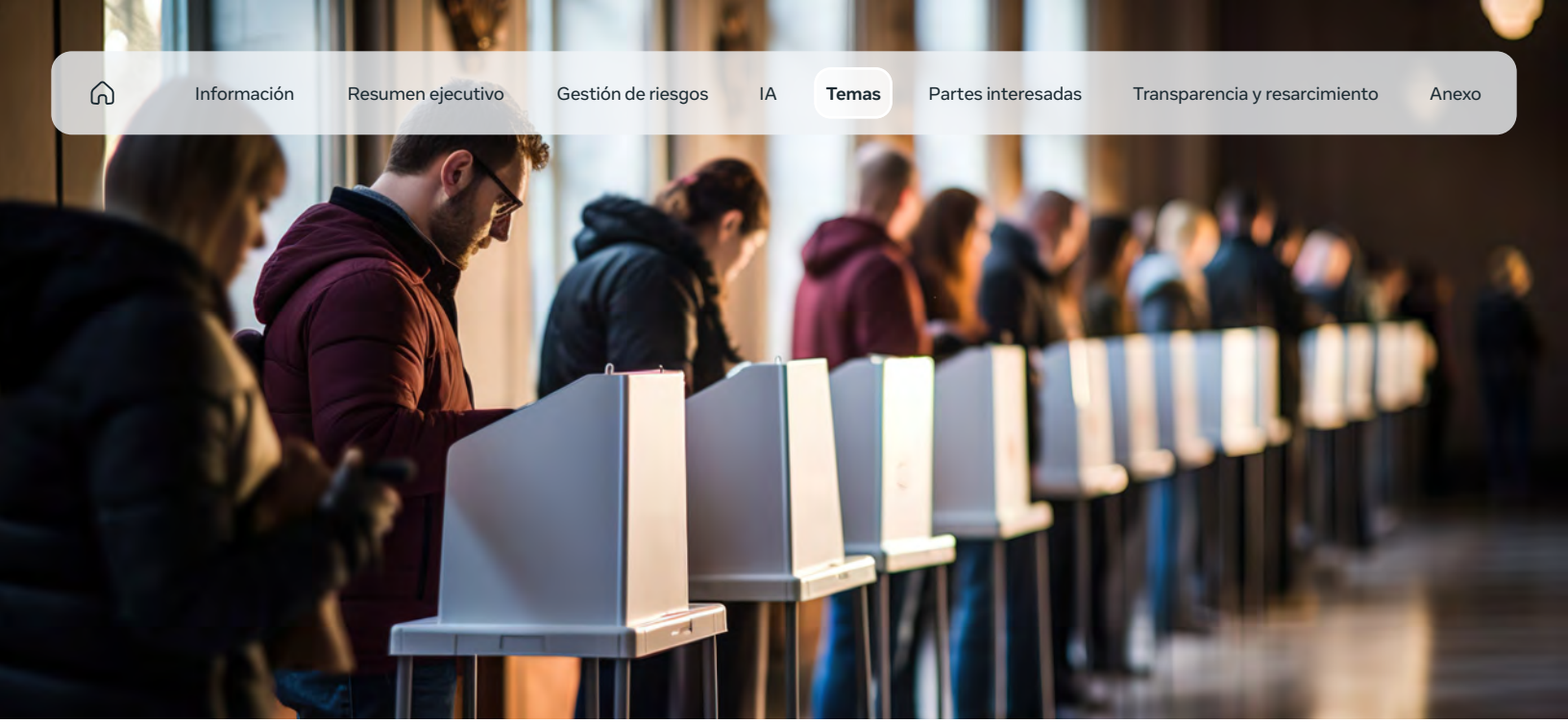
**Gestionar los riesgos de influencia de la IA**

A comienzos del año, a muchas personas les preocupaban los riesgos que podría suponer la IA generativa para unas elecciones justas, incluidos aquellos relacionados con la extensa difusión de deepfakes y las campañas de desinformación con tecnología de IA. Nos preparamos para enfrentar amenazas adversas y la posible interrupción de las elecciones mediante IA y supervisamos este asunto muy de cerca. De los controles que realizamos en nuestros diversos servicios, dedujimos que estos riesgos no se materializaron en gran medida, y el impacto que puedan haber tenido fue leve y de alcance limitado. Por ejemplo, durante el período electoral en un grupo de elecciones importantes, las calificaciones de contenido de IA relacionado con elecciones, política y temas sociales representó menos del 1% de toda la información errónea que se sometió a verificación de datos. Las políticas y los procesos con los que contamos actualmente al parecer bastan para reducir los riesgos que envuelven al contenido de IA generativa.

Durante el año, nuestras iniciativas giraron en torno a detener las operaciones de influencia y aprovechar las colaboraciones con partes internacionales para preservar la integridad de las elecciones.

Supervisamos atentamente el posible uso indebido de la IA generativa por parte de campañas coordinadas que empleaban cuentas falsas. Descubrimos que solo lograron aumentos incrementales en la productividad y la generación de contenido mediante la IA generativa. Estos aumentos incrementales no impidieron que interrumpamos estas operaciones de influencia porque nos centramos en el comportamiento al investigar y eliminar estas campañas, no en el contenido que publican, sea creado con IA o no.

Asimismo, colaboramos con otras partes de nuestro sector para combatir posibles amenazas que surgen del uso de la IA generativa. Por ejemplo, en febrero de 2024, firmamos el Pacto Tecnológico contra el Uso Engañoso de la IA en las Elecciones junto con decenas de otros líderes del sector, en el que nos comprometimos a evitar que el contenido de IA engañoso interfiera con las elecciones mundiales en 2024. En las siguientes páginas, describimos ejemplos de iniciativas relacionadas con la IA específicas de cada país.



### Otras iniciativas de integridad de las elecciones

Además de mitigar los riesgos de la posible influencia de la IA en las elecciones, también intentamos brindar a los votantes herramientas para prevenir la injerencia extranjera, aumentar la seguridad de los candidatos, forjar asociaciones y garantizar la transparencia de los anunciantes.

| | |
|---|---|
| Herramientas para los votantes | El acceso a información confiable y el uso responsable de plataformas online son de especial importancia durante las elecciones. En muchos países, proporcionamos a las personas información del votante y recordatorios el día de las elecciones mediante notificaciones en la app en Facebook e Instagram. Estas funciones permitieron que las personas accedan a información fidedigna de autoridades electorales oficiales sobre el lugar y el momento en los cuales acudir a votar ese día. Por ejemplo, en las elecciones locales realizadas en Brasil, las personas consultaron estas notificaciones aproximadamente 9,7 millones de veces en Facebook e Instagram. Más de 63 millones de personas en Facebook y 118 millones de personas en Instagram vieron el sticker de registro de votantes que las redirigía a información fehaciente sobre las elecciones y las votaciones. |
| Lucha contra la injerencia extranjera | Nuestros equipos de seguridad investigaron y eliminaron redes coordinadas de cuentas, páginas y grupos no auténticos. Asimismo, estimamos que, cada día, nuestro sistema de detección automatizada de cuentas falsas evitó que se creen millones de cuentas falsas. Nuestros equipos eliminaron alrededor de 20 operaciones de influencia encubiertas en todo el mundo, incluido en Oriente Medio, Asia, Europa y los Estados Unidos. Por ejemplo, en Moldavia, eliminamos una red que atacaba a públicos rusoparlantes como parte de nuestras investigaciones sobre presunto comportamiento no auténtico coordinado en la región. |
| Seguridad de los candidatos | Meta también proporcionó medidas de protección más estrictas contra el hackeo, la suplantación de identidad y el acoso destinadas a cuentas de candidatos, funcionarios electos y su personal. Realizamos varias capacitaciones para los candidatos en materia de seguridad que detallaban las pautas disponibles para abordar el acoso en nuestras plataformas y publicamos contenido educativo para que esté ampliamente disponible para todas las personas que participan en las elecciones. |

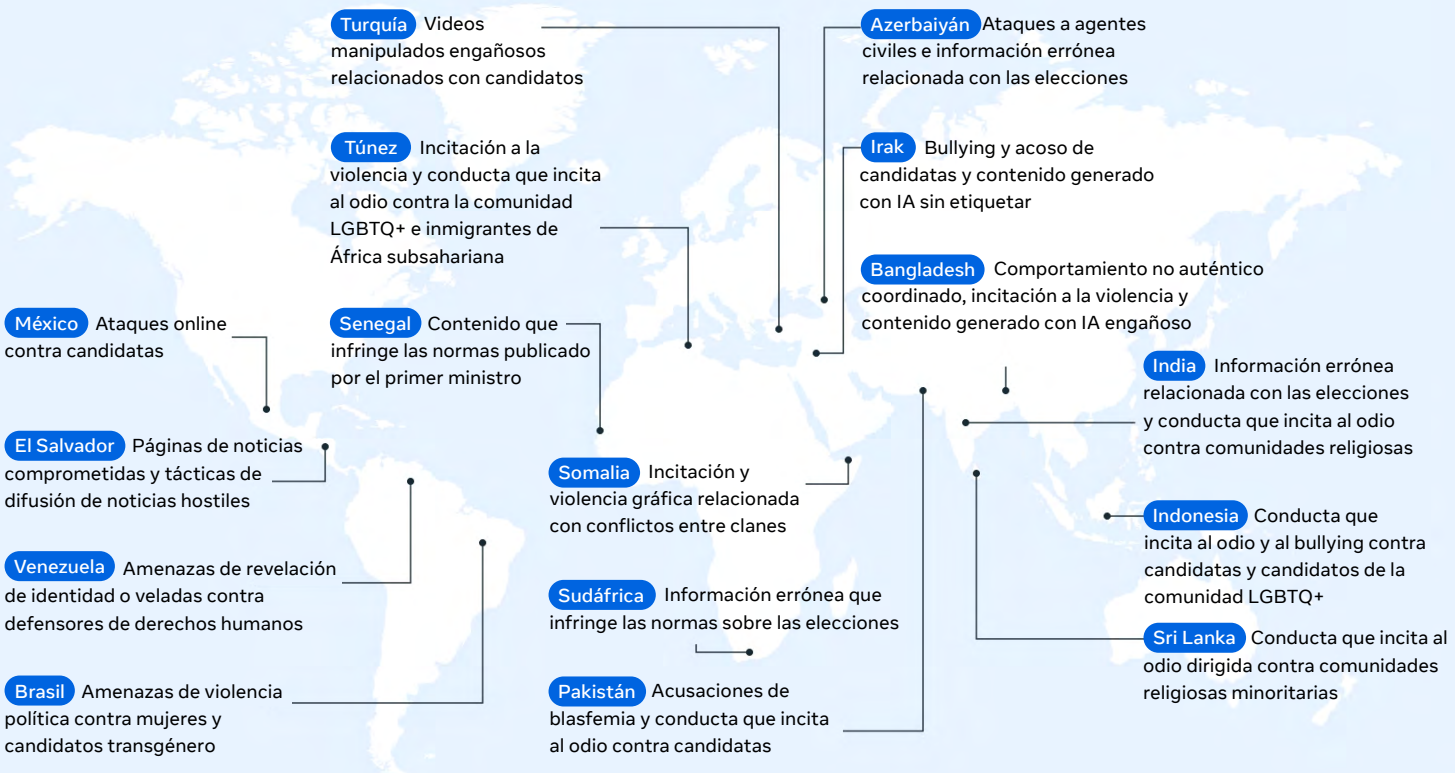
### Vinculación y colaboraciones

Realizamos tareas de vinculación y establecimos canales de comunicación con autoridades gubernamentales y organismos de las fuerzas del orden, de modo que pudieran reportar contenido que potencialmente infringe nuestras Normas comunitarias o la legislación local. También nos asociamos con grupos de la sociedad civil, verificadores de datos y otras empresas tecnológicas para identificar y detener amenazas emergentes y la difusión de información falsa.

### Transparencia de anuncios

Seguimos siendo líderes en el sector en cuanto a la transparencia en anuncios sobre temas sociales, elecciones o política. En la mayoría de los países donde ofrecemos dichos anuncios, los anunciantes se someten a un proceso de autorización y deben incluir un descargo de responsabilidad "Pagado por" en su contenido para publicar sus anuncios. Este descargo de responsabilidad puede incluir información sobre la organización o la persona responsable del anuncio, aunque los requisitos pueden variar según el país. Los anuncios, luego, se almacenan en nuestra biblioteca de anuncios de dominio público. En 2024, agregamos el requisito que exigía que los anunciantes indicaran cuando usan la IA u otras técnicas digitales para crear o alterar un anuncio sobre temas sociales, elecciones o política en ciertos casos.

## Los socios de confianza respaldaron las iniciativas de integridad de las elecciones en 25 países en 2024.



## Preparación para las elecciones de más alto riesgo

Consideramos algunas elecciones de riesgo más alto, lo que exige que nos preparemos mejor, contemos con otros recursos y realicemos una labor a medida. Tenemos en cuenta, por ejemplo, el tipo de elección, el tamaño del país en relación con nuestra base de usuarios, los riesgos de violencia política, los ataques a grupos vulnerables y nuestra capacidad operativa. Otras iniciativas que emprendimos incluyeron establecer esfuerzos de supervisión especializada y medidas temporales de respuesta ante riesgos que se pudieran diseñar y aplicar en diversos países e idiomas.

Abrimos diversos centros de operaciones electorales en todo el mundo para supervisar los problemas que surgían y actuar con rapidez, incluidas en elecciones de más alto riesgo. Puedes consultar información más detallada online sobre nuestras iniciativas en Brasil, Francia, la India, Indonesia, México, Pakistán, Sudáfrica, el Reino Unido, los Estados Unidos y el Parlamento Europeo.

## Ejemplos de elecciones nacionales

Los cuatro ejemplos breves de elecciones en estos países permiten ilustrar cómo trabajamos para gestionar los riesgos electorales en 2024. En cada contexto, comenzamos las preparaciones con, al menos, un año de antelación.

### Estados Unidos

En preparación para las elecciones en los Estados Unidos, nuestras iniciativas se centraron en conectar a las personas con información para el votante confiable, abordar la injerencia extranjera y garantizar la transparencia de los anunciantes.

| | |
|---|---|
| Información para el votante | Durante las elecciones generales de los Estados Unidos en 2024, los recordatorios en la parte superior del feed en Facebook e Instagram recibieron más de 1.000 millones de impresiones. Estos recordatorios incluían información sobre el registro para votar, el voto por correo, el voto anticipado en persona y el voto el día de las elecciones. Las personas hicieron clic en estos recordatorios más de 20 millones de veces para visitar sitios web oficiales del gobierno en busca de más información. |
| Injerencia extranjera | Nos preparamos para abordar la injerencia extranjera online en las elecciones y ampliamos así nuestra aplicación de políticas actual contra entidades de los medios de comunicación controladas por el Estado ruso. Asimismo, seguimos interrumpiendo una de las campañas de influencia encubiertas más grandes y persistentes, denominada Doppelganger. Detuvimos de manera proactiva la gran mayoría de los intentos de la Doppelganger de arremeter contra los Estados Unidos en octubre y noviembre antes de que alguien viera su contenido. |
| Período de restricción publicitaria | Durante la semana final de la campaña electoral, prohibimos la publicación de nuevos anuncios sobre temas sociales, elecciones o política, práctica que mantenemos desde 2020. La lógica que subyace este período de restricción sigue siendo la misma que en años anteriores: en los últimos días de unas elecciones, reconocemos que es posible que no haya tiempo suficiente para impugnar las nuevas afirmaciones hechas en los anuncios. |

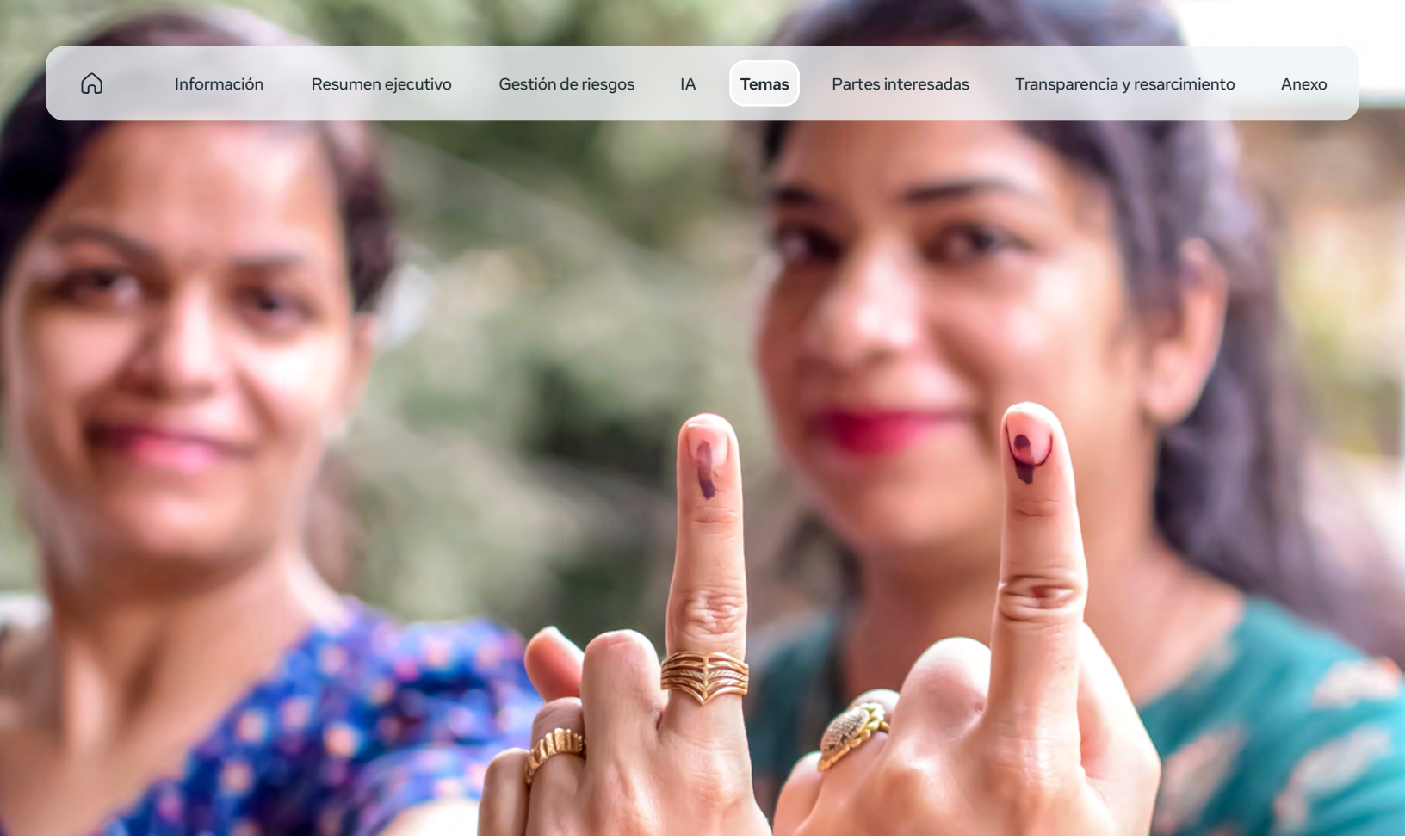
### México

En México, 2024 fue el año con más elecciones de su historia, en el que aproximadamente 90.000 candidatos se postularon para más de 20.000 cargos públicos. La violencia durante las campañas electorales también alcanzó su pico récord. Al menos 37 candidatos fueron asesinados, y se registraron más de 828 ataques sin víctimas fatales. Se postularon más mujeres que durante cualquier otro ciclo electoral en la historia de México, y las candidatas sufrieron altas tasas de violencia y asesinatos debido a su género.

Impulsamos iniciativas similares a aquellas que adoptamos en otros entornos de alto riesgo y nos valimos de los expertos de primera línea de Meta. Eliminamos niveles más altos de contenido infractor que lo habitual antes y en el transcurso de las elecciones. El contenido infractor incluía injerencia en los votantes, venta de votos, contenido de odio y amenazas de acoso y violencia basadas en el género contra las candidatas en Facebook e Instagram.

Para evitar estas interferencias y reducir los riesgos de daño en la vida real, proporcionamos información para el votante y alfabetización mediática fácil de usar como parte de nuestras iniciativas centradas en la seguridad de los candidatos.

| | |
|---|---|
| Seguridad de los candidatos | Registramos a más de 3.000 candidatos, incluidos todos los candidatos a gobernadores y de nivel federal, en nuestro programa de verificación cruzada para prevenir errores en la aplicación de políticas o aplicamos medidas de protección avanzada en sus cuentas. Esto incluyó supervisar posibles amenazas de hackeo. Desarrollamos "Vote Against Violence", una campaña educativa en colaboración con organizaciones sin fines de lucro y grupos de los medios cuyo fin es frenar la violencia de género online. Esta campaña llegó a 1,2 millones de personas en nuestras plataformas y se amplió aún más en otros canales. Las autoridades enviaban solicitudes de eliminación cuando detectaban violencia o amenazas contra los candidatos. |
| Información para el votante | Junto con el National Electoral Institute (INE), lanzamos el chatbot "Inés" en WhatsApp para asistir a los votantes, que respondía preguntas sobre el proceso electoral, por ejemplo, dónde y cómo votar, cómo procesar las tarjetas de identificación del votante y el procedimiento de votación para mexicanos que viven en el extranjero. El día de las elecciones, enviamos recordatorios en Facebook e Instagram, y lanzamos stickers en ambas apps para fomentar el voto. |
| Alfabetización mediática | Para evitar la difusión de información falsa, lanzamos la campaña "Soy Digital" ("We Think Digital") en colaboración con la INE y Movilizatorio, una organización de la sociedad civil. La campaña proporcionaba módulos de aprendizaje accesibles y recursos para consolidar la ciudadanía digital y adquirir competencias de alfabetización en materia de información, incluido cómo proteger la seguridad online. Con la campaña, se llegó a más de 15 millones de personas. También capacitamos a 300 líderes en los distritos electorales que, más tarde, capacitaron a cientos de trabajadores electorales en alfabetización mediática. |



### India

Meta comenzó a prepararse para las elecciones generales de 2024 en la India con 18 meses de antelación. El objetivo era garantizar la integridad de la plataforma y fomentar la educación de los votantes. Adoptamos un enfoque flexible capaz de mantener la continuidad durante un período electoral de 60 días, plazo en el cual se emitieron 640 millones de votos. Nuestros preparativos incluyeron:

| | |
|---|---|
| Educación y concientización de los votantes | Lanzamos la notificación de alerta de votación desde la página de Facebook de la Comisión Electoral de la India, que llegó a 145 millones de personas. Esta comisión implementó también la interfaz de programación de apps (API) de WhatsApp para llevar adelante campañas de recordatorio de votación, con las que se llegó aproximadamente a 400 millones de personas. |
| Medidas para garantizar la integridad de la plataforma | Tomamos medidas para prevenir el uso indebido de nuestras plataformas. Los revisores de contenido analizaron contenido en Facebook, Instagram y Threads en más de 20 idiomas indios y en inglés. Eliminamos cuentas falsas y honramos nuestros compromisos en virtud del Código de ética del voluntario al que, junto con otras empresas de medios sociales, nos adherimos en 2019. |
| Lucha contra la información errónea | Lanzamos una línea de ayuda de verificación de datos específica en WhatsApp con la Misinformation Combat Alliance (MCA) para luchar contra la información errónea generada con IA. Lanzamos una línea de ayuda de WhatsApp con la MCA, que estableció la primera unidad de análisis de deepfakes del mundo para evaluar contenido de audio o video que las personas sospechan que podría ser deepfake. Asimismo, capacitamos a cientos de organismos de las fuerzas del orden junto con la MCA para combatir casos de deepfakes. |

### Elecciones del Parlamento Europeo

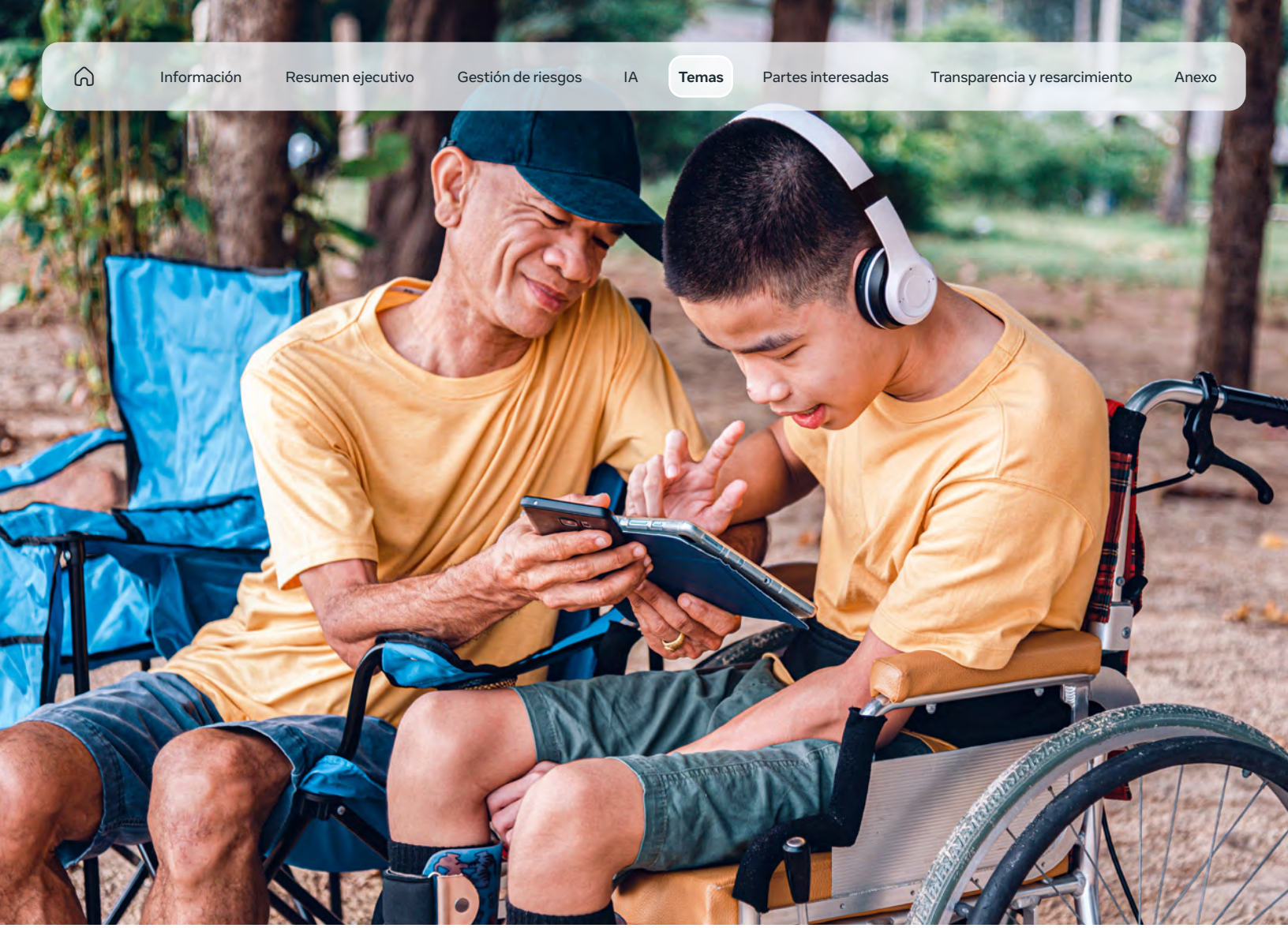
Los preparativos de Meta para las elecciones del Parlamento Europeo se basaron en las lecciones aprendidas de elecciones anteriores realizadas en el mundo, así como en el marco regulatorio conformado en virtud de la Ley de Servicios Digitales y nuestros compromisos en el Código de buenas prácticas en materia de desinformación de la Unión Europea.

Las medidas específicas de la UE que tomamos en cuanto a las elecciones se centraron en lo siguiente:

| | |
|---|---|
| Fomentar la información sobre elecciones y la participación cívica | Proporcionamos datos confiables sobre las elecciones y dirigimos a las personas a información sobre el proceso electoral mediante "unidades de información del votante" en la app e "información del día de las elecciones". Las personas interactuaron con estas notificaciones más de 41 millones de veces en Facebook y más de 58 millones de veces en Instagram. |
| Abordar las operaciones de influencia | Nuestras iniciativas para detener el comportamiento no auténtico coordinado se centraron en amenazas vinculadas específicamente con las elecciones del Parlamento Europeo. Desmantelamos varias redes que acometían a la UE, incluso tomamos medidas varias veces respecto de la red de origen ruso conocida como Doppelganger. |
| Lucha contra la información errónea | Nos asociamos con European Fact-Checking Standards Network para contribuir en la lucha contra contenido multimedia generado con IA y alterado digitalmente, y para llevar a cabo campañas de alfabetización mediática con el fin de aumentar la concientización del público respecto de riesgos relacionados. |
| Lucha contra los riesgos relacionados con el uso indebido de las tecnologías con IA generativa | Como resultado de las políticas y las medidas que aplicamos para abordar el contenido de IA generativa, se incluyó una etiqueta con un descargo de responsabilidad en casi 6.000 anuncios sobre temas sociales, elecciones o política y en más de 5,7 millones de contenidos en Facebook e Instagram en la UE en torno a las elecciones del Parlamento Europeo, con lo que aumentamos la transparencia. |

Consulta más detalles en nuestro Centro de transparencia.

Ir al Centro de transparencia



# Seguridad infantil y juvenil

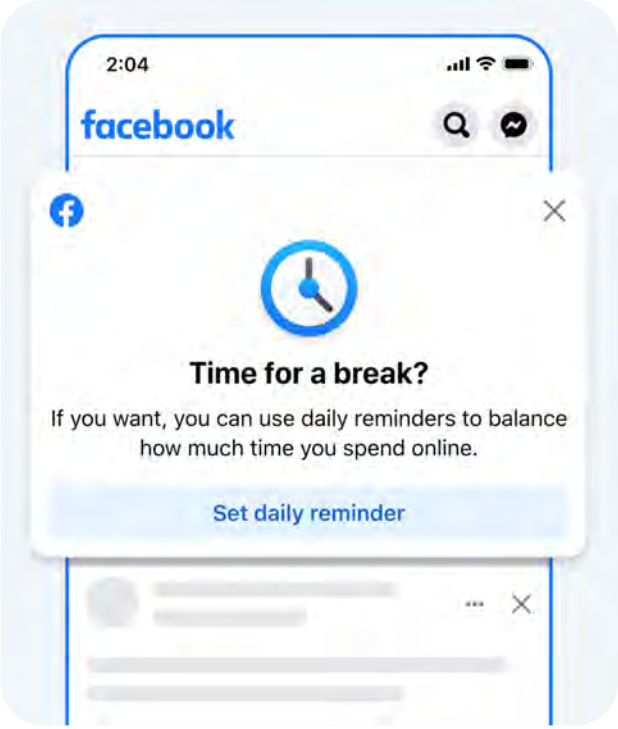
La seguridad de los niños online es una prioridad principal para Meta. Ofrecemos protecciones integradas, así como herramientas para adolescentes y padres cuyo fin es contribuir a la seguridad de los adolescentes en nuestras apps y servicios.

## Protecciones integradas para adolescentes

Para proteger a los adolescentes online, se necesita de la colaboración de varias partes interesadas de todo el mundo, incluidos padres, expertos en infancia, académicos, pares del sector, gobiernos, la sociedad civil, entre otros. Mantenemos nuestro compromiso de proteger a los adolescentes y, al mismo tiempo, brindar un espacio para que ejerzan su libertad de expresión y accedan a información, con sus padres como guía.

En el transcurso de varios años, desarrollamos más de 50 herramientas y recursos en respaldo de adolescentes, padres y tutores, y dedicamos más de una década a elaborar políticas y crear tecnología que aborden el contenido y los comportamientos que infringen nuestras reglas.

En 2024, actualizamos nuestras políticas y el diseño de nuestros productos para ofrecer a los adolescentes una experiencia única diferenciada con mayor claridad. Gracias a estas modificaciones, los adolescentes continúan viendo contenido apropiado para su edad en función de nuestro Marco de protección de los intereses de los niños. Para complementar las protecciones activas actualmente en Instagram, lanzamos las cuentas de adolescente de Instagram en los Estados Unidos, el Reino Unido, Canadá y Australia, que pronto se implementarán en todo el mundo. Las renovadas cuentas de adolescente vienen con protecciones incorporadas que imponen límites respecto de quién puede contactar a los adolescentes y el contenido que pueden ver, así como vías para gestionar el tiempo que pasan en la app. Los cambios también brindan a los adolescentes nuevas formas de explorar sus intereses, guiados por sus padres. Esta nueva experiencia de las cuentas de adolescente de Instagram cumple las recomendaciones de expertos y el principio de las capacidades evolutivas del niño de la Convención sobre los Derechos del Niño de la ONU.




Desarrollamos y lanzamos en todo el mundo un panel de supervisión parental en el que padres y tutores que usan nuestra herramienta de supervisión puedan ver y administrar las cuentas de sus hijos, todo en un solo lugar. Esto permite a padres y tutores configurar controles en su propia cuenta para ver y administrar contactos no deseados o contenido inapropiado, así como establecer límites respecto del tiempo que los menores están expuestos a las pantallas.

Asimismo, realizamos una serie de talleres en los Estados Unidos en torno a nuestro programa Screen Smart, cuyo fin es ayudar a los padres a entablar conversaciones con su familia sobre el uso seguro de dispositivos y obtener información sobre las herramientas de supervisión parental de Meta, de modo que usen los controles y las protecciones más adecuados para ellos.

"Las nuevas cuentas de adolescente de Instagram de Meta son de mucha utilidad, dado que facilitan a los padres que guíen a los adolescentes sin restarles autonomía a quienes ya tienen edad suficiente. La nueva configuración, junto con herramientas y consejos de seguridad y privacidad mejorados, constituyen un avance muy importante".

— Larry Magid, director ejecutivo de ConnectSafely

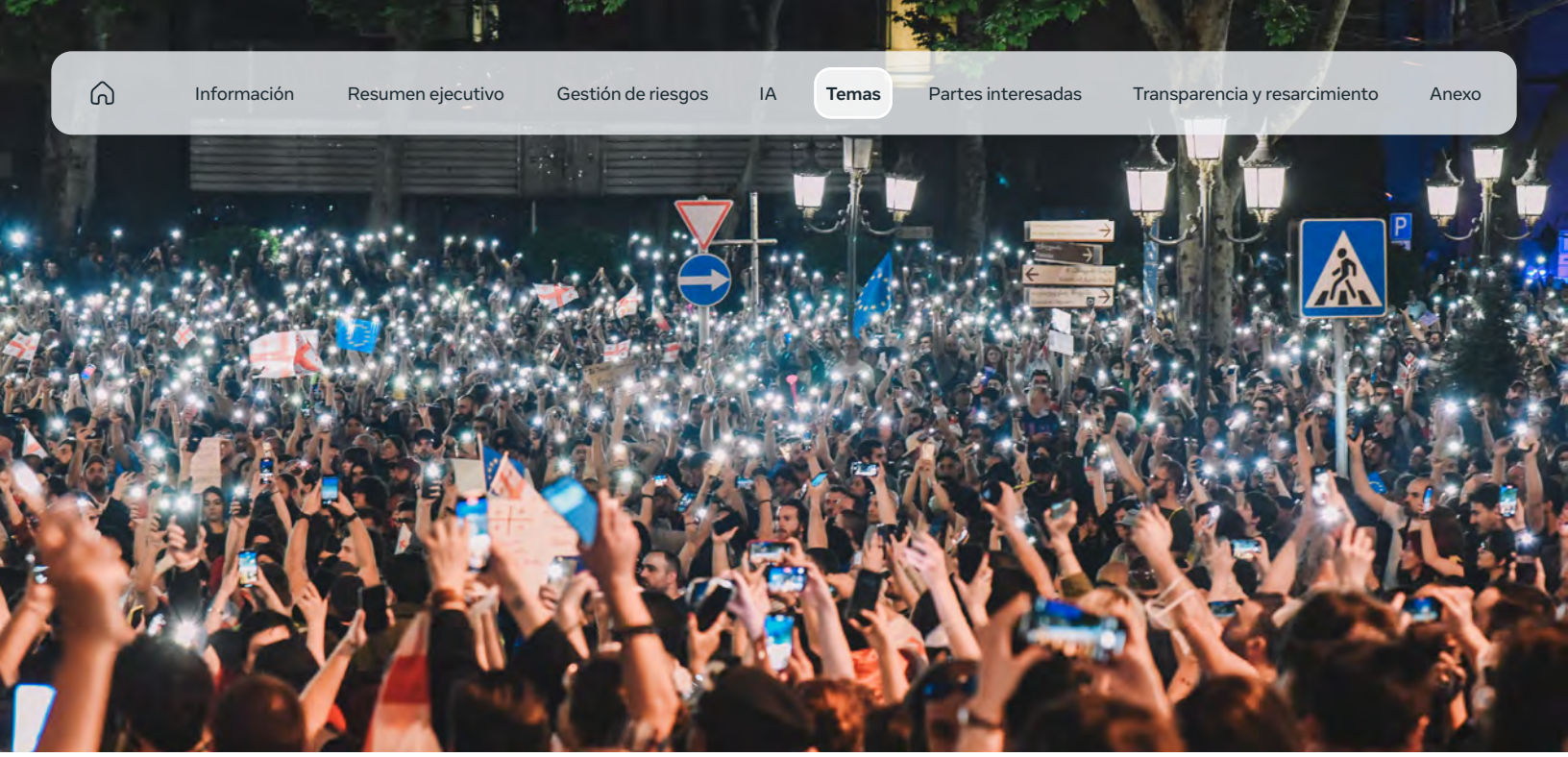
## Lucha contra la sextorsión

La prioridad principal de Meta sigue siendo proteger a los niños de usuarios que intentan hacerles daño. En 2024, seguimos trabajando con el Centro Nacional para Niños Desaparecidos y Explotados (NCMEC) a fin de que el programa de eliminación llegue a más países e idiomas, lo que permite que más adolescentes retomen el control de sus imágenes íntimas. Desarrollamos nuevas herramientas que protegen contra la sextorsión y hacen que resulte más difícil para posibles estafadores y delincuentes encontrar a adolescentes e interactuar con ellos. Asimismo, lanzamos una campaña de concientización, basada en datos aportados por el NCMEC y Thorn, que tiene como objetivo ayudar a los adolescentes a detectar estafas por sextorsión y a los padres, a guiar a sus hijos para que eviten estas estafas. La campaña dirige a adolescentes y padres a consejos de expertos, desarrollados por Thorn y adaptados por Meta, destinados a cualquier persona que solicite ayuda e información relacionadas con la sextorsión.

Somos miembros fundadores de Lantern, un programa que lleva a cabo la Tech Coalition y que permite que nosotros y otras empresas tecnológicas compartamos señales sobre cuentas y comportamientos que infringen las políticas de seguridad de los niños. Compartimos señales específicas de sextorsión con Lantern para consolidar esta importante cooperación entre partes del sector, con el fin de detener las estafas por sextorsión en las diversas plataformas. En 2024, se duplicó la participación en el programa, con un total de 26 empresas inscritas en Lantern.

Consulta la lista completa de herramientas, funciones y recursos de ayuda que ofrecemos a adolescentes y padres aquí.

↗ Leer más



# 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 2024, iniciamos y coordinamos la respuesta ante emergencias en Bangladesh, Corea del Sur, Georgia, Kenia, Nigeria, Nueva Caledonia, el Reino Unido y Venezuela, 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.

El Protocolo de la política de crisis es una herramienta clave que usamos durante un período de crisis. Este guía nuestro uso inmediato de tácticas para mitigar posibles daños en las siguientes áreas:

- Política, como emitir pautas adicionales a los revisores. Un ejemplo es proporcionar pautas para no aplicar faltas a ciertas infracciones de nuestra política de contenido violento y gráfico. El objetivo es evitar imponer penalizaciones excesivas o restringir a los usuarios que intentan generar consciencia respecto del impacto que tiene un conflicto.
- Producto, como cambiar la experiencia de los productos. Como ejemplo, citamos cambiar la configuración, de modo que solo amigos y familiares puedan hacer comentarios en las publicaciones.
- Personas, incluido movilizar recursos para centrarse en problemas concretos.

El Protocolo de la política de crisis nos permite realizar una evaluación offline de situaciones que pueden generar riesgos en la plataforma. Una vez realizada la designación, llevamos a cabo evaluaciones para identificar riesgos en la plataforma y determinar si se deben aplicar otras tácticas. 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.

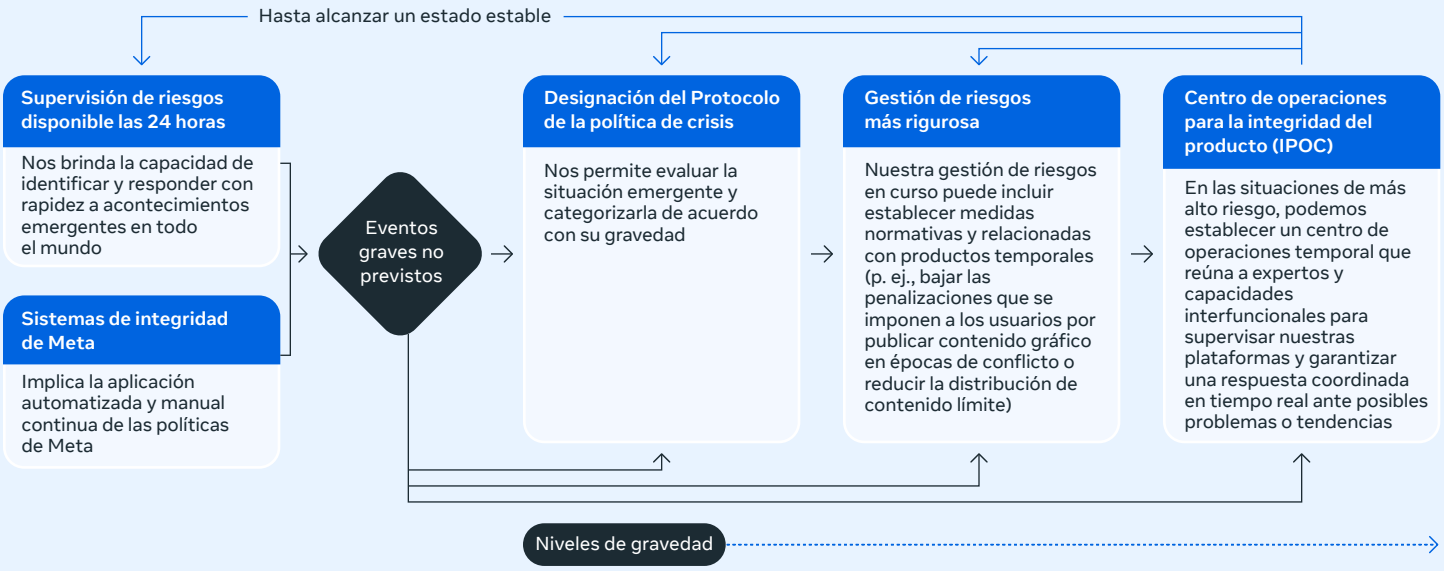
En la próxima página, ilustramos cómo nos preparamos para abordar crisis y conflictos. También brindamos ejemplos para demostrar cómo usamos nuestro Protocolo de la política de crisis y la diversidad geográfica de nuestras iniciativas.

# Preparación para afrontar crisis y conflictos[4]

Nuestro Protocolo de la política de crisis y la labor que realizamos en países en riesgo son las herramientas clave que usamos para prevenir, detectar y mitigar riesgos. Nuestros equipos de productos, políticas y operaciones evalúan las dinámicas locales en evolución para guiar respuestas efectivas y proporcionadas.

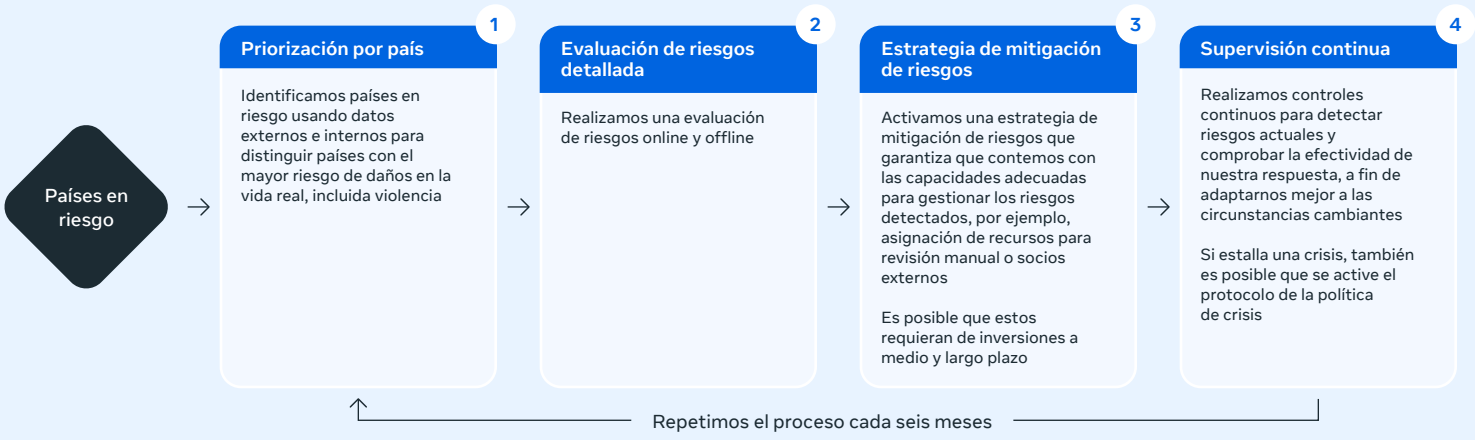
## Respuesta reactiva

### ¿Cómo respondemos rápidamente a acontecimientos graves imprevistos?




## Medidas a largo plazo

### ¿Cómo tomamos medidas a largo plazo para mitigar los riesgos de conflicto?




[4] Nuestra labor en materia de crisis incluye muchas situaciones, incluidos conflictos, violencia intracomunitaria, disturbios civiles, manifestaciones masivas y desastres ambientales, así como ataques terroristas y otros ataques criminales, en todo el mundo.

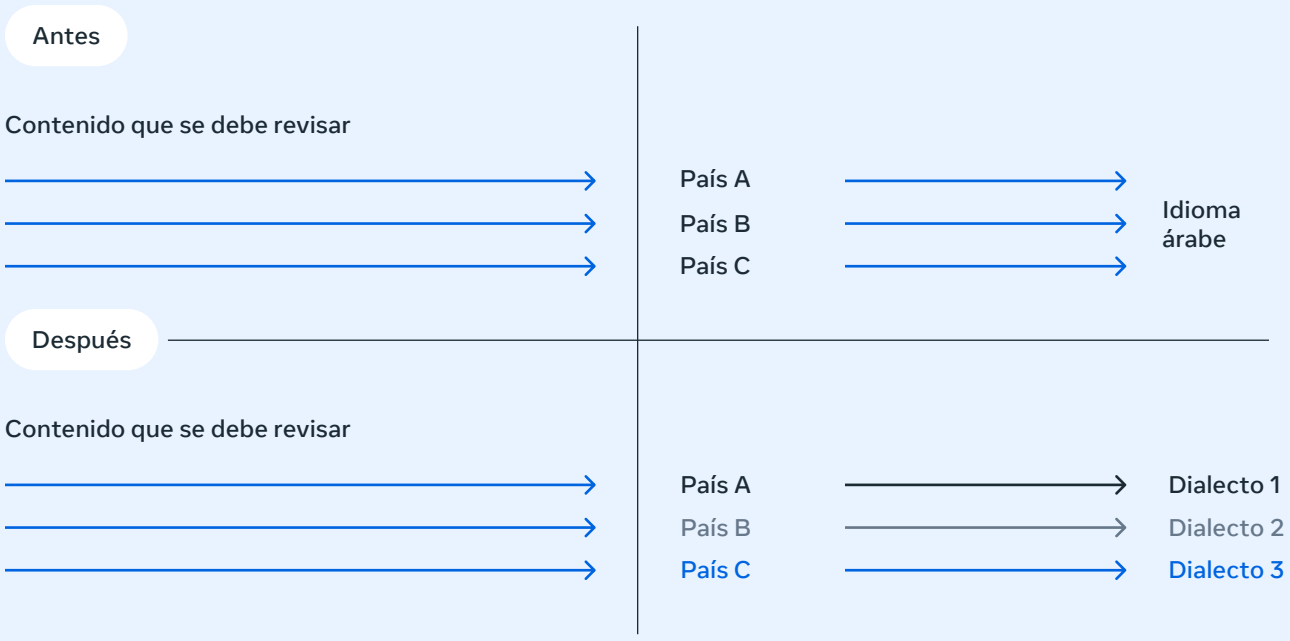
## Sudán

En 2024, se intensificó aún más el conflicto en Sudán entre las Fuerzas Armadas de Sudán (FAS) y las Fuerzas de Apoyo Rápido (FAR), lo que exacerbó la inestabilidad y la crisis humanitaria del país. El volumen de contenido infractor, incluida la violencia e incitación, daño coordinado, explotación de personas, y personas y organizaciones peligrosas, aumentó en comparación con los niveles previos al conflicto y se mantuvo alto a lo largo del año.

Para reducir la prevalencia de contenido infractor, continuamos nuestra labor sobre la base de las medidas que tomamos en 2023, con el Protocolo de la política de crisis como guía. Como el conflicto no cesaba, implementamos medidas temporales y desarrollamos mitigaciones a largo plazo para abordar los riesgos de que el volumen de contenido infractor continuara siendo alto.

Una de estas mitigaciones a largo plazo fue diseñar, crear y lanzar un sistema que identifica dialectos árabes y deriva el contenido de forma prioritaria a revisores con más probabilidades de entender los matices lingüísticos y el contexto local. El sistema anterior identificaba el árabe como idioma único y remitía el contenido a moderadores con capacidad de revisarlo. El nuevo sistema puede detectar el dialecto particular del árabe empleado y derivar el contenido al revisor con más probabilidades de entenderlo. En Sudán, este cambio supuso mayores volúmenes de contenido revisado con más precisión, lo que redujo errores de aplicación de políticas. Este trabajo se basó en los resultados de la diligencia debida en materia de derechos humanos en Israel y Palestina y se desarrolló a partir de ellos.

### Canal para matices lingüísticos



Durante 2024, ambas partes implicadas en el conflicto revelaban cada vez más las identidades de los prisioneros de guerra online. Hacerlo aumentaba el riesgo de daños en el mundo real y perjudicaba la protección de la dignidad y la seguridad de estas personas, conforme lo indica el Convenio de Ginebra relativo al trato debido a los prisioneros de guerra. Tomando como referente una recomendación que el Consejo asesor de contenido hizo en 2023 y reiteró en 2024 en el caso Cautivo en video de las Fuerzas de Apoyo Rápido de Sudán, reconocemos también que cierto contenido sobre prisioneros de guerra podría ser de interés público, por ejemplo, para generar consciencia sobre los posibles abusos a los derechos humanos o ayudar a localizar a estas personas. En consecuencia, Meta proporcionó pautas a los revisores de contenido con respecto a los prisioneros de guerra en la Política de organización de actos dañinos y promoción de la delincuencia, de modo que pudieran abordar de mejor manera el contenido potencialmente infractor en la región a gran escala.

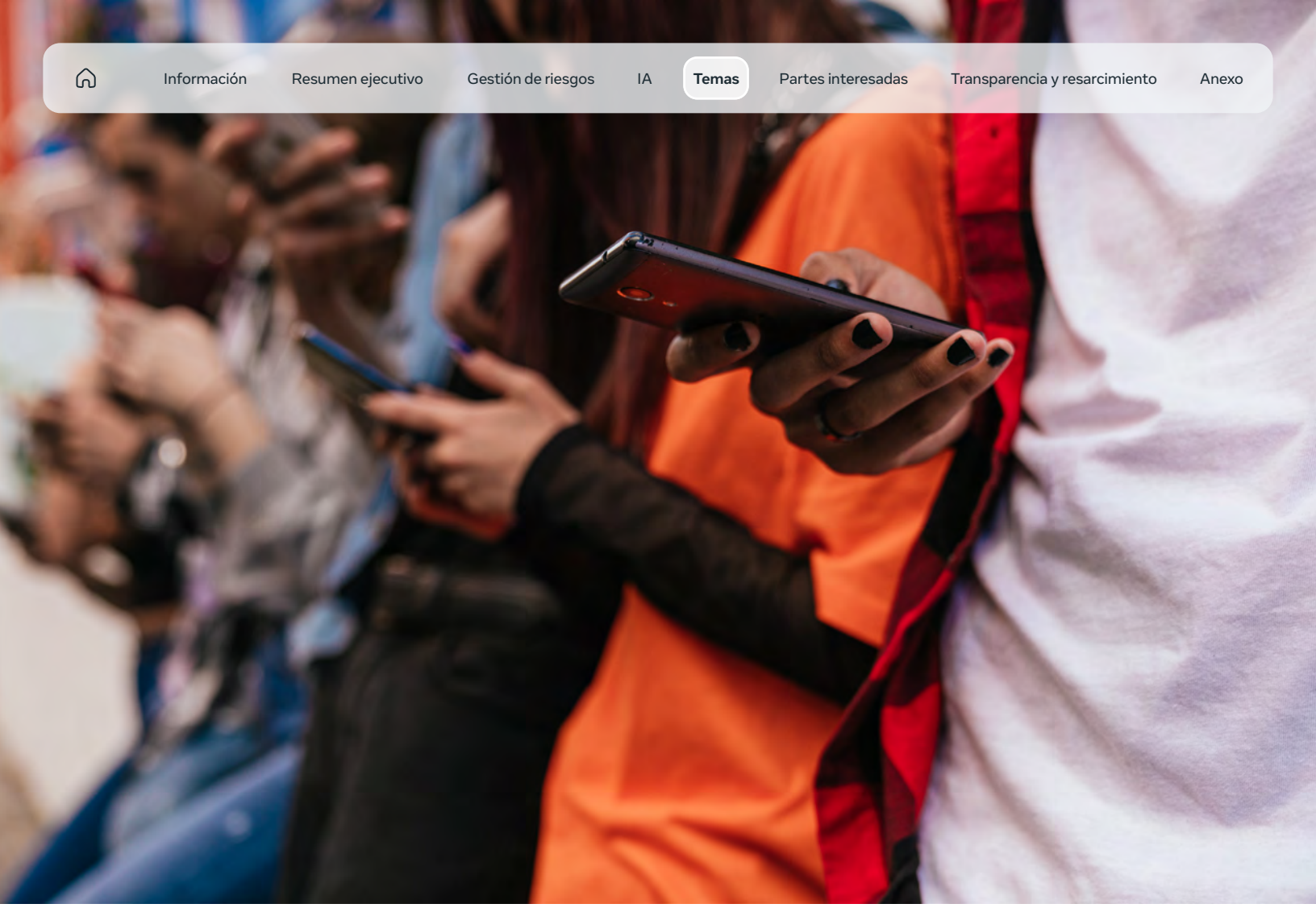
Leer el Informe del Consejo asesor de contenido del primer semestre de 2024

Consultar el caso del Consejo asesor de contenido

Los socios de confianza desempeñaron un papel fundamental, ya que brindaron información decisiva sobre los desarrollos locales y el contenido potencialmente infractor relacionado con el conflicto. Gracias a esta información valiosa, se pudieron aplicar las políticas de Meta relevantes, incluidas las de conducta que incita al odio, bullying y acoso, explotación de personas, así como la designación de afirmaciones potencialmente dañinas revisadas previamente en virtud de la Política de información errónea y daño y esto, en última instancia, contribuyó a un entorno online más seguro. Conforme a la Política de explotación de personas, pudimos identificar los riesgos potenciales relacionados con imágenes y reclutamiento de soldados menores de edad, eliminamos ese contenido y redujimos su prevalencia.

Realizamos sesiones de capacitación con defensores de los derechos humanos, periodistas, y organizaciones nacionales y de la diáspora para que estos a su vez pudieran capacitar a usuarios sudaneses, incluidos inmigrantes y refugiados. Estas sesiones se centraron en las políticas de contenido y la seguridad digital, y en mejorar su presencia en las plataformas de Meta.

Los conflictos armados desencadenan el desplazamiento masivo de personas, así que, en Sudán, nos centramos en detectar la posible explotación, incluidos el tráfico ilegal de personas, la explotación sexual de mujeres y niñas, y el matrimonio forzado. Eliminamos más de 19.100 publicaciones en grupos en los que se ofrecían servicios de tráfico ilegal de personas, así como contenido que glorificaba el matrimonio forzado. Colaborar con la diáspora siguió siendo una estrategia importante para detectar tendencias de contenido y brindar claridad en cuanto a la supervisión en el país. Esto incluyó más de 30 colaboraciones únicas para proteger a los usuarios en nuestras plataformas, incluidas iniciativas para identificar términos y frases nuevos y emergentes relacionados con el lenguaje que incita al odio y el tráfico ilegal de personas. Gracias a estas estadísticas, logramos detectar con mayor eficacia contenido que infringía nuestras políticas y responder en consonancia.



## Oriente Medio

El conflicto en Oriente Medio siguió siendo una prioridad para Meta. En 2024, pusimos el foco en los riesgos derivados de la violencia constante en Israel y Gaza, a medida que la guerra se extendía por toda la región, y otras partes procedentes de ella se involucraban más y se intensificaba el conflicto. Trabajamos para garantizar que nuestras plataformas sean un espacio donde ejercer la libertad de expresión y, al mismo tiempo, prevenir la difusión de contenido que incita al terrorismo, la violencia y otros daños en el mundo real.

Mantuvimos el enfoque principal de 2023, que incluía mantener la designación del ataque de Hamás del 7 de octubre de 2023 como atentado terrorista en virtud de nuestra política de personas y organizaciones peligrosas y abordar el contenido infractor conforme a nuestras políticas. Interrumpimos los cambios temporales en los productos que introdujimos en 2023.

Inmediatamente después del atentado terrorista del 7 de octubre, Meta designó la violencia y el posterior conflicto al nivel más alto de nuestro Protocolo de la política de crisis e implementó medidas de respuesta ante emergencias inmediatas. Entre estas medidas, se estableció un equipo interfuncional especializado disponible las 24 horas, así como medidas temporales normativas y de producto. Basamos nuestro enfoque en los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas, que es pilar de nuestra Política corporativa de derechos humanos, y en nuestra debida diligencia en 2022. Se pueden consultar detalles de nuestra respuesta en nuestro Informe sobre derechos humanos de 2023 y en publicaciones de Newsroom.

A lo largo de 2024, seguimos trabajando en conjunto con una variedad de agentes del Gobierno, la sociedad civil y otras partes en Israel y los países árabes de Oriente Medio, pero también a nivel mundial, para demostrar transparencia y capacidad de respuesta. También respondimos a varios casos del Consejo asesor de contenido.

Asimismo, seguimos implementando las recomendaciones del informe de diligencia debida en materia de derechos humanos de 2022. Emitimos informes del progreso de Meta en el período comprendido entre el 30 de junio de 2023 y el 30 de junio de 2024, incluido un aumento de nuestros recursos de moderación de contenido en idioma hebreo y una actualización a nuestra política de personas y organizaciones peligrosas cuyo fin es permitir más discursos sociales y políticos en determinadas situaciones. Tomamos esta medida en respuesta a los comentarios que indicaban que nuestra política de personas y organizaciones peligrosas se aplicaba con demasiada frecuencia a contenido como noticias, debates neutrales de acontecimientos actuales o incluso repudio contra grupos terroristas y grupos de odio. Seguimos prohibiendo el contenido que exalta o respalda a personas u organizaciones peligrosas, o los actos o las misiones de violencia que emprenden.

Entre el 30 de junio de 2024 y el 30 de junio de 2025, lanzamos un sistema que detecta el contenido y prioriza derivarlo a moderadores que es más probable que comprendan ese dialecto árabe particular. Renovamos también el canal de escalamiento a socios de confianza, lo que condujo a una respuesta más rápida a los escalamientos. Se puede consultar nuestro progreso durante este período en nuestra actualización final de diciembre de 2025: Diligencia debida en materia de derechos humanos de Israel y Palestina.

↗ Progreso logrado en 2023 ↗ Progreso logrado en 2024

## Bangladesh

Nuestros preparativos para las elecciones de 2024 permitieron que nos anticipáramos a varios riesgos durante el período del informe y pudiéramos abordarlos. Nuestro objetivo era proteger a los usuarios y, al mismo tiempo, respaldar su capacidad para votar y expresar su opinión. Gracias a estos preparativos electorales, logramos responder, a mitad de año, a las manifestaciones estudiantiles, las represiones violentas y el posterior cambio de gobierno.

Dada la gravedad de los disturbios, implementamos nuestro Protocolo de la política de crisis. Identificamos riesgos de forma proactiva, incluidos lenguaje que incita al odio, incitación en contra de comunidades religiosas minoritarias, información errónea y comportamiento no auténtico coordinado. Implementamos medidas de mitigación, como el uso de nuestra política de lugar de alto riesgo temporal, la colaboración con socios de confianza y la red de verificación de datos independiente, así como la incorporación de protecciones más rigurosas para las cuentas de defensores de los derechos humanos.

Nuestras otras medidas incluían:

- Establecer señales de detección precisas para identificar aumentos de contenido relacionado con infracciones al que se podría aplicar políticas en tiempo real, como la de violencia gráfica y conducta que incita al odio.
- Emplear herramientas y técnicas, incluida detección mediante IA, para identificar y aplicar políticas a contenido infractor y búsquedas de palabras clave.
- Designar otras afirmaciones potencialmente dañinas revisadas previamente en virtud de la Política de información errónea y daño.

No cumplimos con las solicitudes del Gobierno de eliminar el contenido sobre las manifestaciones si no eran coherentes con las normas de derechos humanos internacionales. Esto se hizo de acuerdo con nuestros compromisos como miembros de la Global Network Initiative y nuestra Política corporativa de derechos humanos.

## Georgia

Implementamos el Protocolo de la política de crisis dos veces en Georgia en 2024. En marzo de 2024, lo implementamos primero luego de una serie de manifestaciones masivas en oposición a la Ley de Transparencia de la Influencia Extranjera. Lo reactivamos en diciembre de 2024 luego de las elecciones nacionales, cuando se llevaron a cabo otras manifestaciones multitudinarias que derivaron en una escalada de violencia por parte de la policía y otras fuerzas de seguridad.

Implementar el Protocolo de la política de crisis permitió a nuestro equipo mejorar sus iniciativas de mitigación, abordar los picos de contenido infractor y riesgos de violencia física acrecentados, y proteger a los defensores de derechos humanos. Auditamos la lista de insultos, palabras que históricamente se emplean para atacar a determinados grupos, para identificar y gestionar contenido de odio en nuestras plataformas. Eliminamos cuentas falsas que se diseñaron para manipular la opinión pública o distribuir contenido potencialmente dañino. Asimismo, interrumpimos una red de comportamiento no auténtico coordinado en Georgia, así como otras cuentas no auténticas.

Durante los períodos de crisis, colaboramos con organizaciones de la sociedad civil, verificadores de datos y socios de confianza, quienes nos ayudaron a entender la situación en curso y facilitaron el intercambio de información con la sociedad civil más general y la oposición en Georgia. Los socios de confianza proporcionan señales e información crítica respecto del contenido infractor que arremete contra grupos de la oposición. Además, en colaboración con ellos, ayudamos a estos grupos a entender mejor lo que constituye contenido infractor conforme a nuestras Normas comunitarias. Asimismo, trabajamos en conjunto con socios de la sociedad civil para identificar a los defensores de derechos humanos en riesgo, a fin de garantizar que se apliquen medidas de protección más rigurosas a su cuenta.

# Ciberseguridad

Nuestras políticas de seguridad son importantes para los derechos a la libertad de expresión, el acceso a la información y la privacidad de los usuarios, entre otros. Seguimos trabajando en toda la empresa para identificar y contrarrestar amenazas adversas contra las plataformas, incluidas operaciones de influencia, ciberespionaje, vigilancia, fraude y estafas. Un elemento importante de nuestra labor en materia de seguridad es interrumpir redes adversas que se involucran en actividades maliciosas.

En 2024, eliminamos 20 redes de comportamiento no auténtico coordinado en Oriente Medio, Asia, Europa y los Estados Unidos por infringir nuestra política de comportamiento no auténtico coordinado. La labor de estas redes es manipular el debate público para conseguir un objetivo estratégico mediante el uso de cuentas falsas o tácticas engañosas. Supervisamos y detenemos los intentos de reconstitución en nuestras plataformas de redes que ya eliminamos, compartimos información con el público mediante nuestros informes de amenazas y trabajamos para desarrollar estadísticas a partir de investigaciones y basar en ellas nuestros sistemas de detección y el diseño de productos, de modo que sean más resilientes.

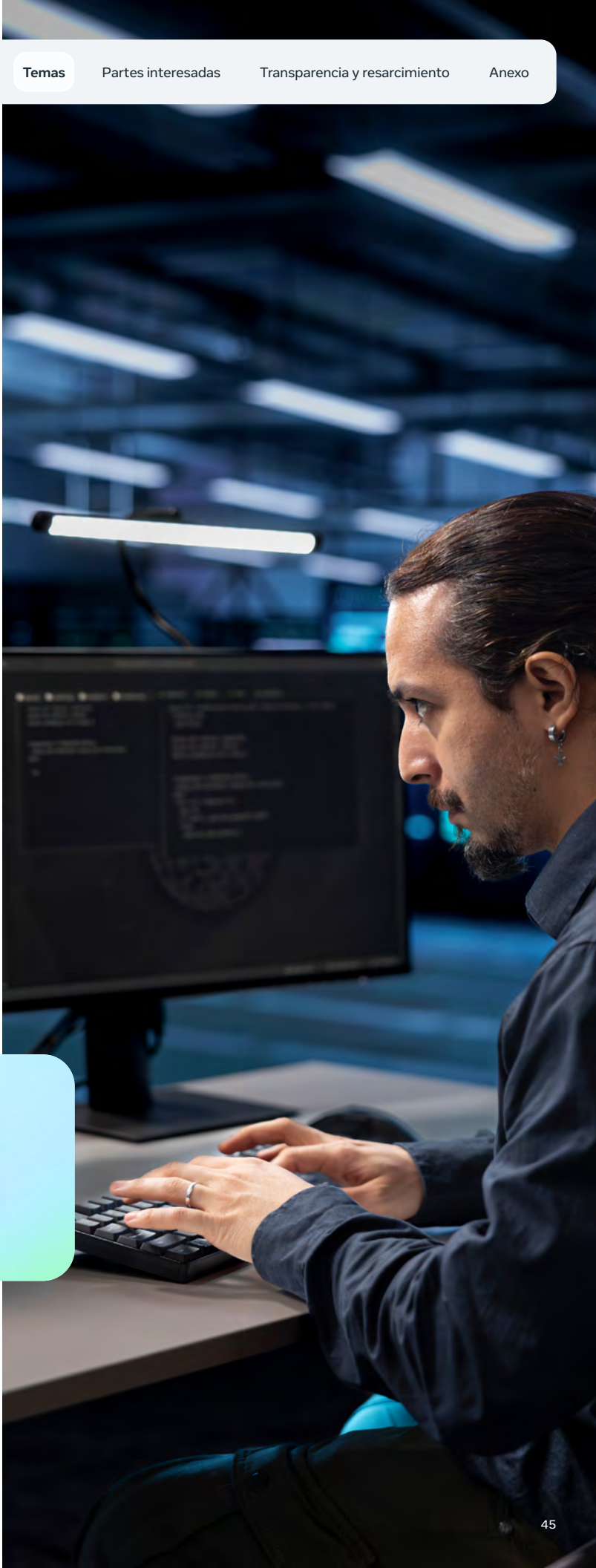
Seguimos detectando y eliminando redes de comportamiento no auténtico coordinado que atacaban a grupos religiosos o étnicos específicos o se hacían pasar por ellos. En 2024, uno de muchos ejemplos fue una red que se originó en Bangladesh, que eliminamos debido a comportamiento no auténtico coordinado. Esta usaba cuentas falsas para publicar contenido y administrar páginas con el fin de atacar a públicos nacionales. La red se hacía pasar por entidades de noticias ficticias y usaba el nombre de organizaciones de noticias reales para difundir contenido contra el Partido Nacionalista de Bangladés y en respaldo del partido en el poder. La operación se vinculaba con personas asociadas a la Liga Awami y a una organización sin fines de lucro.

Otro ejemplo es una red de China cuyo objetivo era la comunidad sij y que, mediante cuentas falsas y comprometidas, se hacía pasar por sij y fomentaba un movimiento activista ficticio llamado Operación K, que incitaba a organizar manifestaciones sij, incluido en Nueva Zelanda y Australia. La operación usaba imágenes y publicaciones generadas con IA, en inglés e hindi, sobre inundaciones en la región del Punjab, la comunidad sij de todo el mundo, el movimiento independentista Khalistan, el asesinato de Hardeep Singh Nijjar y la crítica del Gobierno indio.

Leer más

Como parte de nuestras iniciativas de aplicación de políticas contra empresas de spyware, interrumpimos y eliminamos la actividad de Paragon Solutions, un proveedor de spyware que tomaba como objetivo a diversos usuarios de WhatsApp, incluidos periodistas y miembros de la sociedad civil. Nos comunicamos con los usuarios de WhatsApp que podrían haber resultado afectados y les brindamos recursos para que aprendan a protegerse. Asimismo, les proporcionamos información sobre The Citizen Lab en la Universidad de Toronto, que brinda otros recursos a miembros de la sociedad civil. En 2024, fuimos signatarios fundadores del Pall Mall Memorandum, una iniciativa multinacional para frenar el uso indebido de spyware.

En diciembre de 2024, un juez federal de los Estados Unidos declaró culpable a NSO Group de infringir leyes estatales y federales, e incumplir las Condiciones del servicio de WhatsApp. Esta fue la primera vez que la ley estadounidense declara culpable a una empresa de spyware. Meta y WhatsApp demandaron en 2019 a NSO Group, que había accedido a los servidores de WhatsApp sin autorización para instalar el spyware Pegasus en los dispositivos móviles de más de 1.400 usuarios de WhatsApp, incluidos periodistas, activistas de derechos humanos, disidentes políticos, entre otros.



**20**

redes de comportamiento no auténtico coordinado eliminadas

# 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 2024, colaboramos con una amplia gama de partes, incluidos miembros de la sociedad civil, académicos, grupos de expertos, especialistas en derechos humanos y autoridades reguladoras. Las preguntas normativas clave incluían nuestro enfoque respecto de una inteligencia artificial responsable y la integridad de las elecciones, así como nuestras señales de designación de personas y organizaciones peligrosas, y eventos violentos.

Por ejemplo, evaluamos si nuestra política relacionada con el término "sionista" era apropiada, realizamos consultas con 145 partes interesadas de la sociedad civil y el sector académico de todo el mundo. Las participantes incluían científicos políticos, historiadores, académicos del derecho, grupos de derechos civiles y digitales, defensores de la libertad de expresión y expertos en derechos humanos. También participaron partes interesadas que incluían organizaciones sin fines de lucro de nuestro Programa de socios de confianza, así como una amplia gama de comunidades de la diáspora que representaban diversos puntos de vista.

En 2024, creamos un equipo de trabajo de voz y expresión con organizaciones de la sociedad civil locales en las regiones de África subsahariana, Oriente Medio y África septentrional para conocer sus inquietudes con las propuestas legislativas del Reino de Arabia Saudita, Jordania, Nigeria y Senegal, entre otros países. Durante estas sesiones, exploramos cómo proteger el acceso a nuestras plataformas y, al mismo tiempo, gestionar las restricciones de contenido en función de la legislación local y nuestros compromisos con la Global Network Initiative de defender la libertad de expresión y la privacidad de los usuarios.

También realizamos un piloto de un programa de organismos de derechos humanos, que incluían instituciones de derechos humanos nacionales de Etiopía, Ghana, Kenia, Nigeria y Sudáfrica, y se centraban en cómo Meta aborda el contenido potencialmente dañino y la regulación del contenido online.

Asimismo, realizamos talleres de respuesta ante conflictos en Etiopía, Palestina, Somalia, Sudán y Túnez. Capacitamos a defensores de los derechos humanos y periodistas de países en períodos electorales para dotarlos de las herramientas necesarias para proteger su presencia digital.

Mediante el trabajo realizado en nuestro programa Open Loop India y Open Loop Sprint, colaboramos con empresas, legisladores y expertos en IA para producir estadísticas sobre el rol que desempeña la colaboración con partes interesadas en todo el ciclo de vida y la cadena de valor de la IA.

## Nuestro enfoque respecto de la colaboración con partes interesadas

- **Incorporar una amplia gama de perspectivas y conocimiento**: desentrañar información importante y colaborar con expertos en la materia de todas las regiones para obtener opiniones generales diversas y conocer los matices locales.
- **Proporcionar transparencia**: analizar desafíos y mejoras con partes externas.
- **Crear un ciclo de comentarios**: mostrar cómo evolucionan nuestras políticas con el tiempo.
- **Generar confianza**: reflejar legitimidad en nuestras políticas y su aplicación.

# 464

partes interesadas de **34** países contribuyeron a **seis** flujos de trabajo del foro de políticas.

# 121

partes interesadas contribuyeron con otras labores de desarrollo de Meta.

# +100

partes interesadas participaron en sesiones informativas sobre las elecciones, y se produjeron **siete** newsletter electorales.

# +290

periodistas, defensores de derechos humanos y activistas recibieron capacitación.

# Ciclo de desarrollo de políticas de Meta

**Revisión constante**

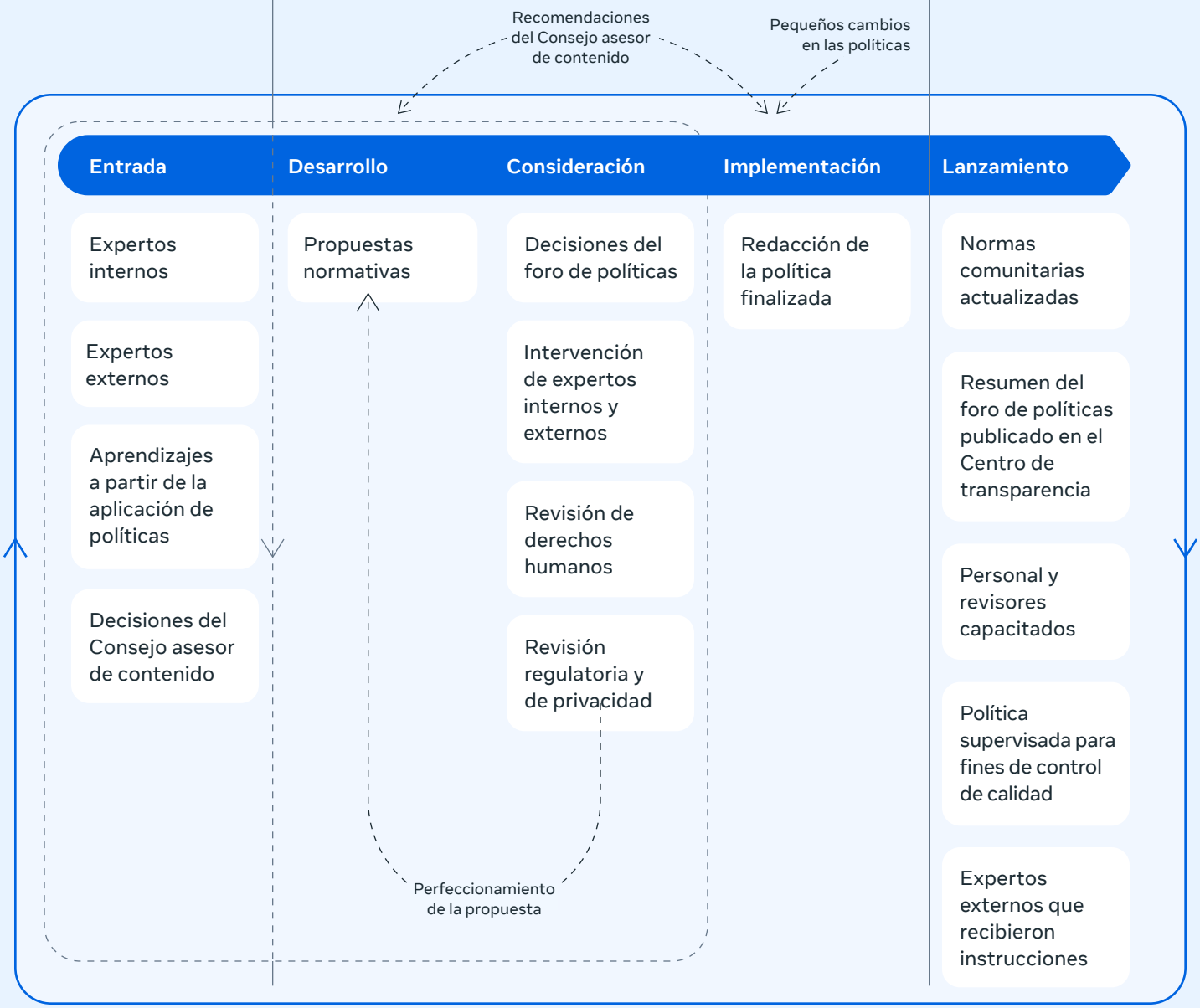
Revisamos continuamente nuestras políticas sobre la base de los aportes que obtenemos de diversas fuentes.

**Desarrollo**

Las propuestas se someten a un proceso de desarrollo riguroso para garantizar que respeten los principios, funcionen y sean fáciles de explicar.

**Lanzamiento**

Los sistemas de aplicación de políticas se actualizan y las políticas "se ponen a disposición" en nuestros servicios.



## Foro de políticas

Pretendemos elaborar políticas que respeten los derechos humanos y apuntamos a adoptar perspectivas diversas, en las que se escuchen y reflejen varias opiniones y creencias. El foro de políticas es una reunión frecuente en la que expertos en la materia debaten posibles cambios a las Normas comunitarias y las Normas de publicidad. Estas reuniones implican proponer nuevas políticas o modificar las existentes, seguir un proceso de desarrollo normativo que incluya una extensa colaboración con partes interesadas de todo el mundo y una revisión tanto de investigaciones externas como internas.

**Realizamos seis foros de políticas en 2024:**

1. "Sionista" como término indirecto que representa una conducta de odio
2. Eventos violentos infractores
3. Contenido comercial con posibles riesgos para la salud y la seguridad
4. Eliminación de imágenes sensibles
5. Contenido relacionado con trastornos alimenticios
6. Condolencias por personas peligrosas designadas

# Foros de la comunidad

Los foros de la comunidad de Meta se arraigan en la gobernanza deliberativa y se diseñaron para aprovechar la participación pública en problemas en los que hay intercambios contrapuestos, pero no hay respuestas claras. Nuestro enfoque permite que personas ajenas a la empresa tangan mayor espacio para intervenir en nuestra toma de decisiones y que podamos observar la evolución que puede tener la opinión pública en el futuro.

En 2024, Meta realizó un foro de la comunidad en asociación con Deliberative Democracy Lab de Stanford, que giraba en torno a los principios que los usuarios desean que sustenten el desarrollo de agentes de IA. El foro contó con la participación de alrededor de 1.000 personas de la India, Nigeria, Arabia Saudita, Sudáfrica y Turquía. Hay un informe detallado disponible aquí.

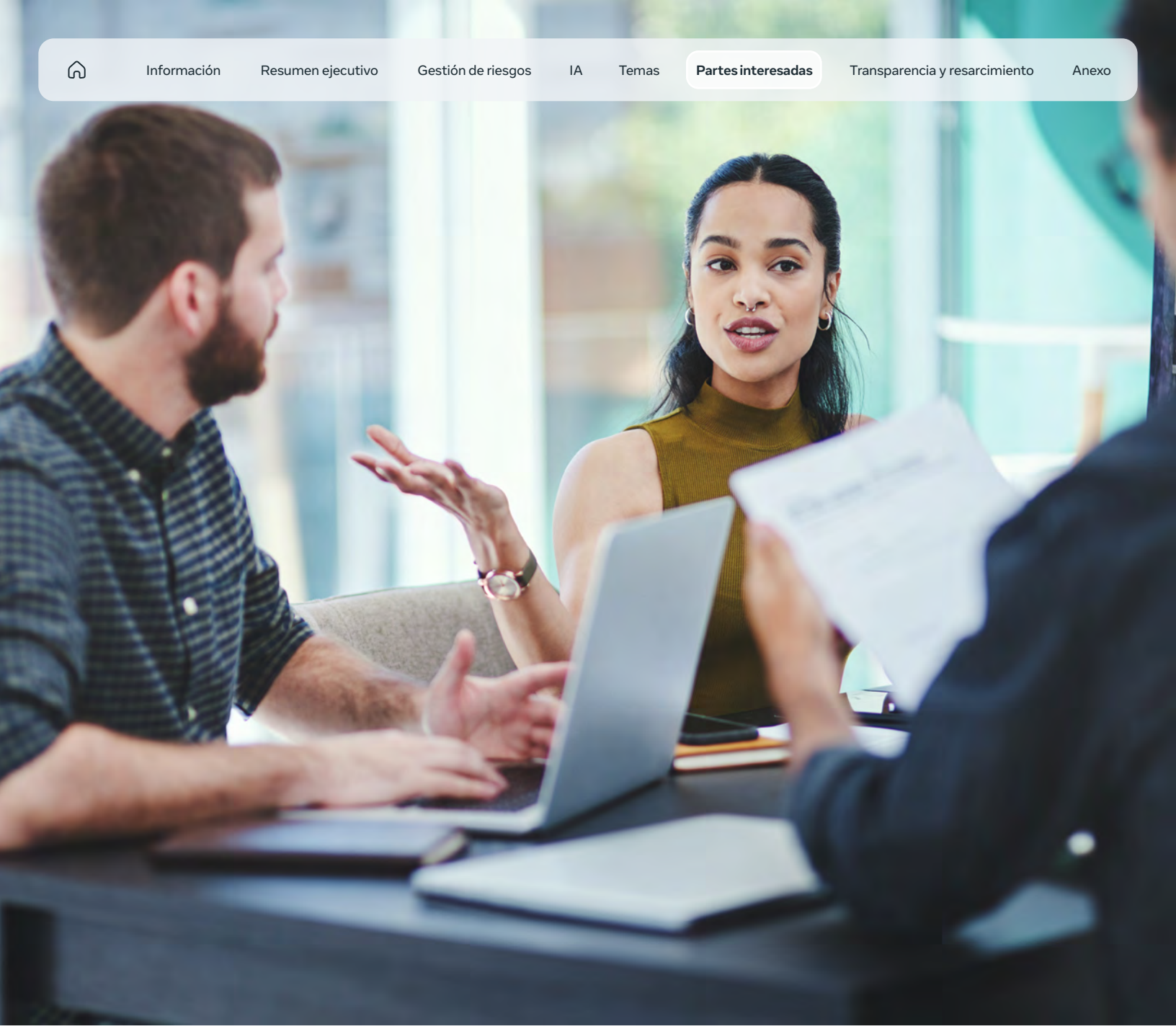
Como parte del foro, los participantes pudieron oír por vía directa a expertos en la materia, reflexionar con los demás y ofrecer a Meta comentarios valiosos. El método de reflexión les permitía abordar las tensiones inherentes que surgen al brindar experiencias personalizadas, ponderar el valor de la personalización con el costo que supone, como la recopilación y el almacenamiento de datos.

## Nuestro enfoque respecto de los controles de usuario y las experiencias personalizadas

Basamos en los hallazgos obtenidos nuestro enfoque respecto de los controles y las experiencias personalizadas que brindamos a los usuarios con los agentes de IA. Estos incluían:

- Los participantes respaldaron a los agentes de IA recordando sus conversaciones anteriores para personalizar su experiencia, en especial si había transparencia y controles disponibles para los usuarios.
- Los participantes apoyaron más a los agentes de IA personalizados conforme a la cultura o la región que a aquellos estandarizados.
- Los participantes favorecieron a los agentes de IA que se asemejan a personas, que pueden responder a manifestaciones emocionales.

Asimismo, comenzamos un piloto para que el público exprese qué considera que contribuye a un modelo de IA relevante, y para poder desarrollar conjuntos de datos de preferencia en función de estos comentarios y proporcionar datos a los desarrolladores en formato de código abierto. El resultado sería una colección de fácil acceso de conjuntos de datos para que nuestro macromodelo lingüístico Llama resulte más relevante y útil en diferentes contextos culturales.



## Socios de confianza

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.

Los socios de confianza son importantes aliados a la hora de identificar infracciones graves a nuestras Normas comunitarias y fueron particularmente útiles durante 2024, año electoral. Proporcionaron estadísticas e identificaron contenido dañino en países en circunstancias de mucho disturbio, como Bangladesh, Brasil, Costa de Marfil, Francia, Grecia, India, Indonesia, Kenia, México, Nigeria, Pakistán, Región del Kurdistán, República Democrática del Congo, Senegal, Siria, Sudáfrica y Venezuela, entre otros países y regiones.

En 2024, eliminamos más de 100.000 contenidos infractores gracias al Programa de socios de confianza.

Los socios de confianza brindan estadísticas sobre tendencias de contenido relacionado con las elecciones para basar en ellas las iniciativas de integridad, ayudarnos a detectar y eliminar contenido infractor, e identificar a usuarios de nuestra plataforma que corren un alto riesgo para ofrecerles otras protecciones. Los socios de confianza lograron identificar con efectividad los aumentos de lenguaje hostil dirigido contra comunidades marginalizadas y los ataques a periodistas y defensores de derechos humanos, así como el uso indebido de contenido creado con IA.

Para abordar el riesgo de conducta que incita al odio, eliminamos los insultos designados. Trabajamos en colaboración con nuestros socios de confianza para entender mejor el contexto en el que se emplean insultos, de modo que pudiésemos aplicar nuestras políticas con mayor precisión.

Recurrimos a más de 40 socios de confianza de 20 países para que nos brindaran información en la cual basar los procesos de desarrollo de políticas y productos en lo que respecta a eliminación de imágenes sensibles, explotación de personas, personas y organizaciones peligrosas y señales de designación de personas, "sionista" como término indirecto que representa una conducta de odio, chatbots de IA y más.

En respuesta a la recomendación del Consejo asesor de contenido, Meta evaluó la puntualidad y efectividad de las respuestas ante contenido reportado mediante el Programa de socios de confianza. Durante un período de dos años comprendido entre el segundo trimestre de 2022 y el cuarto trimestre de 2024, Meta realizó importantes mejoras en cuanto al tiempo de respuesta ante el contenido reportado mediante este programa.

Gracias al dinero invertido en capacitación, sistemas de aplicación de políticas optimizados y nuevas herramientas, mejoramos el volumen de reportes y la eficiencia de la revisión en 2024.

### Resultados a nivel internacional

En todo el mundo, el Programa de socios de confianza recibió más de **11.800** contenidos reportados en el segundo trimestre de 2022, cifra que ascendió a **49.200** en el segundo trimestre de 2024, es decir, **se multiplicó por cuatro**.

## Crecimiento mundial del canal de socios de confianza en el plazo de dos años

T2 de 2022 a T2 de 2024

**+4 veces**
más contenido reportado mediante el programa de socios de confianza

**+12 puntos**
de porcentaje de casos resueltos en un plazo de cinco días a partir del escalamiento

**+15%**
de reducción eficiente del tiempo de respuesta mediano en días

**+15 veces**
más contenido reportado para someter a revisión normativa

Se incluyen ejemplos del impacto que tuvo el programa de socios de confianza en Pakistán, Siria y Venezuela en las siguientes páginas.

CASO DE ÉXITO

# Informes elaborados a partir de estadísticas de Siria

Durante el período posterior a la caída del régimen de Assad en diciembre de 2024, los socios de confianza desempeñaron un papel fundamental, ya que se encargaron de elaborar informes y analizar los desarrollos locales, proporcionar estadísticas sobre tendencias de contenido locales y escalar infracciones graves.

Tomando como base los informes elaborados por los socios de confianza, dotados de su conocimiento local, desarrollamos las iniciativas de respuesta ante emergencias de Meta y logramos aplicar nuestras políticas, así como mitigar el riesgo de un modo más oportuno y eficiente. Los socios de confianza plantearon inquietudes sobre los riesgos actuales y las declaraciones de afiliación con el régimen derrocado que atentan contra minorías étnicas y religiosas, incluidos los alauitas, los cristianos y los kurdos, y señalaron el aumento de diferentes facciones extremistas dentro del antiguo ejército sirio. Con estas estadísticas respaldamos las iniciativas que emprendemos para mitigar el riesgo de que haya personas y organizaciones peligrosas en nuestras plataformas y de que ocurran ataques físicos debido a características personales.

CASO DE ÉXITO

# Mitigación de riesgos para los agentes civiles en Venezuela

En el período previo a las elecciones del 28 de julio de 2024 en Venezuela, trabajamos en colaboración con nuestros socios de confianza y entablamos nuevas relaciones con organizaciones de la sociedad civil a fin de prepararnos para afrontar los riesgos relacionados con las elecciones y aumentar los reportes de contenido infractor.

En este período, se desataron manifestaciones, que el Gobierno intentó sofocar con represión, incluidos detenciones masivas y arrestos coordinados de oponentes políticos. Nuestros socios de confianza brindaron estadísticas indispensables sobre los desarrollos in situ. Estos reportaron contenido dañino, entre el cual se lanzaban amenazas veladas y se develaba la identidad de los manifestantes y de los seguidores de la oposición, lo que los exponía al riesgo de detenciones arbitrarias y daños físicos. Asimismo, los socios de confianza reportaron ataques a cuentas de agentes civiles, como periodistas, miembros de la oposición y defensores de derechos humanos, entre otros.

Gracias a esta información valiosa, logramos realizar una detección proactiva, así como ofrecer protección avanzada a estas cuentas y evitar errores de aplicación de políticas con ayuda de la verificación cruzada. Estas medidas respaldaron la difusión de noticias y la participación cívica en un entorno represivo.

CASO DE ÉXITO

# Los socios de confianza abordan las acusaciones de blasfemia y el lenguaje hostil en Pakistán

En Pakistán, los socios de confianza fueron clave, ya que nos alertaron respecto de contenido potencialmente dañino cuyo objetivo eran comunidades marginadas, incluidas minorías religiosas y de género.

Durante el período electoral de febrero de 2024, los socios de confianza reportaron contenido relacionado con las elecciones que incluía lenguaje hostil dirigido contra candidatos políticos y acusaciones de blasfemia que podrían derivar en incitación. En Pakistán, las acusaciones de blasfemia pueden desencadenar acciones judiciales y violencia física.

Gracias a estos reportes, logramos eliminar contenido relacionado con acusaciones de blasfemia en virtud de nuestra política de organización de actos dañinos y promoción de la delincuencia.

Los socios de confianza también proporcionaron señales y estadísticas durante otros momentos críticos, como los estallidos de violencia sectaria. Su labor nos permitió responder de inmediato, eliminar el contenido infractor en nuestras plataformas y fortalecer nuestra detección y aplicación de políticas.

## Colaboración con partes interesadas en Pakistán

Meta emprendió una serie de colaboraciones en Pakistán con varias partes interesadas gubernamentales y no gubernamentales como parte de la diligencia debida en materia de derechos humanos. Algunos puntos destacados incluyen:

Un debate abierto sobre la seguridad de los jóvenes online, organizado conjuntamente con el Ministerio de Derechos Humanos, la Comisión Nacional de los Derechos del Niño, la Comisión Nacional de Derechos Humanos y la Fundación de Derechos Digitales. Hablamos sobre las cuentas de adolescente y el lanzamiento del portal Take It Down en urdu para los usuarios pakistaníes.

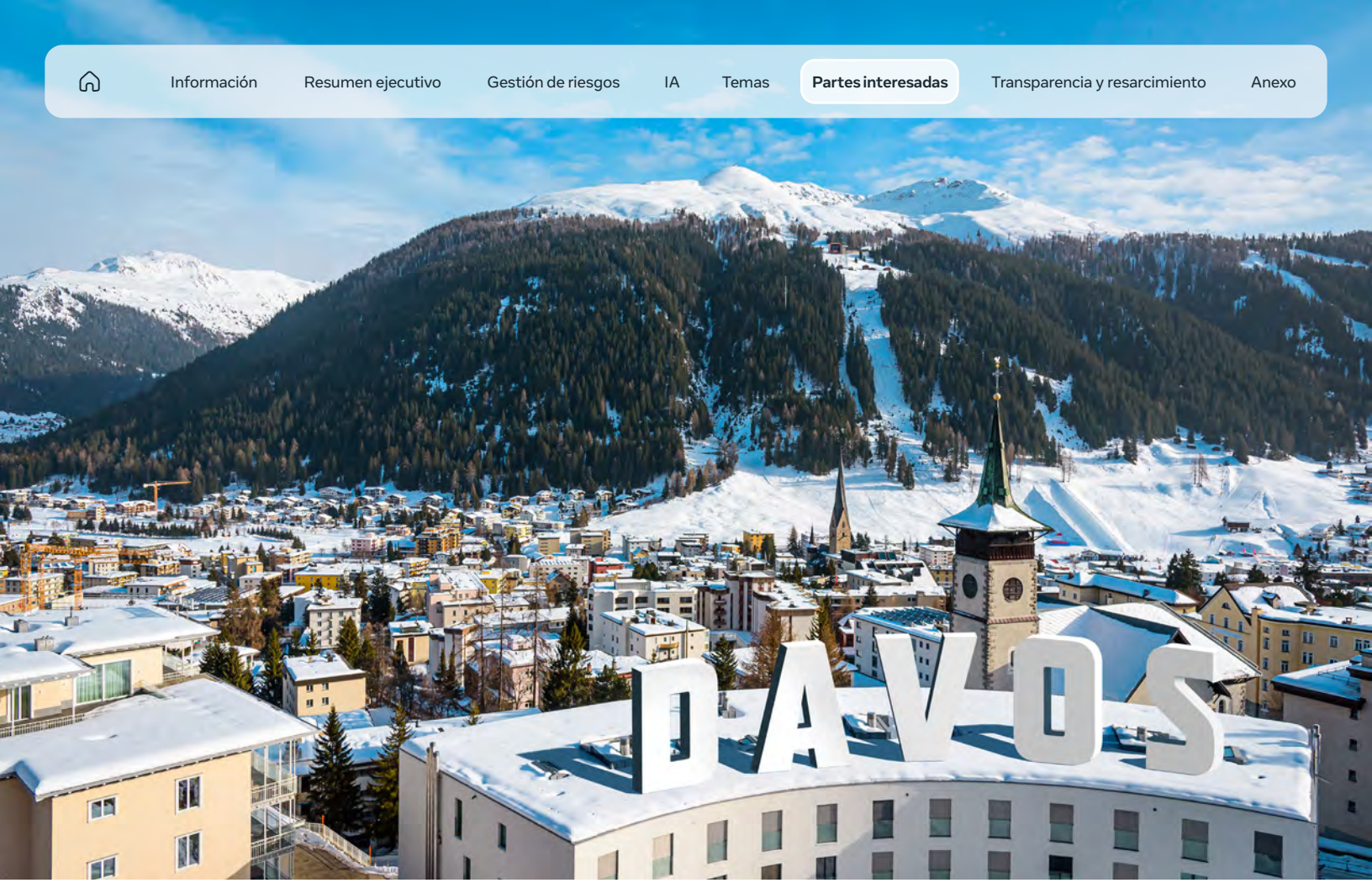
---

Trabajo conjunto con un grupo diverso de defensores de derechos humanos para recopilar estadísticas sobre las interrupciones de internet y explorar las posibles vías de colaboración en relación con la defensa de causas. Estos brindaron información valiosa sobre el impacto del "firewall" y la autorización de redes privadas virtuales por parte del Gobierno.

---

Un debate abierto con organizaciones de la sociedad civil para evaluar en profundidad los compromisos que Meta asumió con los derechos humanos clave y el trabajo de nuestros equipos de derechos humanos. Esto incluyó un debate sobre cómo responder a situaciones sin suspender el servicio de internet o limitar el rendimiento de las plataformas de medios sociales, incluida nuestra familia de apps.

Cada interlocutor, en cada evento, mencionó el impacto de realizar acusaciones adversas y frívolas de blasfemia contra los usuarios. Transmitimos tranquilidad al señalar la política de riesgo de exposición de Meta y el trabajo que realizamos constantemente para proteger a las personas que son objeto de estas acusaciones.




## Organizaciones internacionales

En 2024, los Estados miembros de las Naciones Unidas negociaron y adoptaron el Pacto Digital Mundial (GDC), un marco integral para la gobernanza internacional de las tecnologías digitales y la IA. Trabajamos con los Estados miembros de la ONU, organismos de la ONU y coaliciones del sector para finalizar el texto del GDC. Nuestra labor apuntaba a amparar la libertad de expresión y, al mismo tiempo, crear un futuro digital más seguro, inclusivo y abierto para todos.

Meta también tuvo otras participaciones dentro del sistema de la ONU a lo largo del año. Nuestra tarea incluía contribuir con el informe Tecnologías Digitales, Derechos del Niño y Bienestar de UNICEF para realizar la debida diligencia en el sector tecnológico y con la UNESCO en cuanto a la gobernanza de la desinformación en las plataformas digitales. Asimismo, apoyamos a la UNESCO con una interfaz de traducción basada en el modelo de IA Meta No Language Left Behind (NLLB) para proporcionar traducciones de calidad a 200 idiomas, entre ellos, idiomas minoritarios como asturiano, luganda, maorí, suajili y urdu, a fin de fomentar la diversidad lingüística y el acceso a la información.

Meta siguió trabajando en estrecha colaboración con la Oficina del Alto Comisionado para los Derechos Humanos (OHCHR). Nos reunimos regularmente con personal del OHCHR y tuvimos una participación activa en el Proyecto B-Tech, que proporciona pautas fidedignas y recursos para implementar los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas en el sector tecnológico y su Comunidad de Práctica, un espacio que propicia el diálogo confidencial con otras empresas tecnológicas. Participamos activamente en debates continuos sobre IA y normas de derechos humanos. Asimismo, formamos parte del Foro de la ONU sobre las Empresas y los Derechos Humanos de 2024 e hicimos presentaciones en paneles sobre el "lenguaje que incita al odio online" y "cómo proteger la libertad de prensa".

Meta se sumó a los debates normativos paralelos a la Cumbre del Futuro y la Asamblea General de la ONU n.º 79. Entre los temas que se trataron estaban el rol de la IA en la gobernanza mundial, el empoderamiento de los creadores digitales, la innovación impulsada por la diáspora y el impacto de las leyes sobre ciberdelito y contenido en la libertad de expresión. Participamos también en debates sobre la protección que se brinda a los defensores de derechos humanos y el uso de medios sociales para transmitir información que salva vidas durante crisis humanitarias.

Asimismo, consultamos con los Procedimientos Especiales del Consejo de Derechos Humanos de la ONU (expertos en derechos humanos independientes), incluidos Relatores Especiales de la ONU sobre la libertad de expresión y defensores de derechos humanos, entre otros.

A lo largo del año, Meta colaboró con el G7, el G20, la UNESCO and la OECD en flujos de trabajo relacionados con la inclusión y la gobernanza de la IA. Asimismo, nos sumamos a conversaciones con los gobiernos en las que debatimos la importancia de la integridad de la información. Seguimos participando de la Coalición Mundial para la Seguridad Digital del Foro Económico Mundial, de la que se desprendió la publicación del informe The Intervention Journey: A Roadmap to Effective Digital Safety Measures.

Asimismo, participamos activamente y colaboramos con varias partes interesadas en diversos foros, incluidos la Cumbre Mundial para Erradicar el Odio, el Foro sobre la Libertad en Internet en África (FIFAfrica), el Foro Mundial de Internet para la Lucha contra el Terrorismo, el Foro de Gobernanza de Internet (FGI), RightsCon, la Conferencia sobre tecnología contra la trata de personas y la Comisión de la Condición Jurídica y Social de la Mujer de la ONU.

Mediante nuestra membresía en la Global Network Initiative y nuestra participación en la Alianza para la Confianza y la Seguridad Digital, Meta asistió al Foro de Participación de las Partes Interesadas sobre Derechos y Riesgos en Europa, que nos permitió basar en la información obtenida nuestras evaluaciones de riesgos sistémicos en virtud de la Ley de Servicios Digitales.

# Transparencia y resarcimiento

El Consejo asesor de contenido, un organismo independiente, nos ayudó 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 reglas vinculantes sobre si 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.

### Dónde obtener más información sobre el impacto del Consejo asesor de contenido

En 2024, pasamos de elaborar informes trimestrales a semestrales sobre casos que Meta remitió al Consejo asesor de contenido y actualizaciones de nuestro progreso en cuanto a la implementación de sus recomendaciones. Asimismo, lanzamos una página del Centro de transparencia que realiza un seguimiento del impacto que tienen las recomendaciones del Consejo asesor de contenido. Esto se suma a nuestra página de recomendaciones del Consejo asesor de contenido, donde detallamos las recomendaciones relacionadas con un caso que este nos envía, nuestro nivel de compromiso con ella y el estado de implementación.

# Acciones relacionadas con las recomendaciones del Consejo asesor de contenido en 2024

| Recomendaciones que emitió el Consejo asesor de contenido | Evaluación o implementación de Meta en curso[5] | Recomendaciones implementadas[5] |
|---|---|---|
| 48 (66 en 2023) | 70 (69 en 2023) | 41 (61 en 2023) |

En 2024, 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, incluidos el derecho a la libertad de expresión, a la salud y a la no discriminación, entre otros. Aquí ofrecemos algunos ejemplos ilustrativos de nuestras medidas en respuesta a las decisiones que tomó el Consejo asesor de contenido en 2024. Consulta los detalles en los informes semestrales de Meta sobre el Consejo asesor de contenido.

## Entre los ejemplos de las decisiones tomadas por el Consejo asesor de contenido en 2024 se incluyen las siguientes:

El Consejo asesor de contenido anuló las decisiones de Meta de eliminar tres publicaciones de Facebook en las que se mostraban imágenes del ataque terrorista de Moscú en marzo de 2024 y exigió que el contenido se restaurara con pantallas de advertencia "Marcar como perturbador". El Consejo asesor de contenido determinó que, si bien las publicaciones infringían las políticas de Meta, ya que mostraban el momento en que ocurrieron los ataques designados contra víctimas visibles, eliminarlas no fue una medida coherente con las responsabilidades de la empresa con los derechos humanos.

El Consejo asesor de contenido ratificó la decisión de Meta de eliminar un video en el que aparecía un político pakistaní dando un discurso con texto que afirmaba que esta persona estaba "cruzando todos los límites de la lealtad" y usando el término "kufr" que sugiere una blasfemia. Esta medida se tomó debido al riesgo de daño en la vida real.

[5] Algunas evaluaciones o implementaciones en curso o recomendaciones implementadas en su totalidad incluyen recomendaciones de años anteriores (consulta nuestro Informe sobre derechos humanos de 2023 para obtener más información).

## Entre los ejemplos de las medidas que tomamos a partir de las recomendaciones del Consejo asesor de contenido se incluyen los siguientes:

Tras una serie de recomendaciones (p. ej., aquí) sobre IA, realizamos cambios a la manera en que gestionamos el contenido generado con IA, incluidas etiquetas y políticas actualizadas, como nuestra política sobre información errónea.

---

Siguiendo las recomendaciones del Consejo asesor de contenido sobre la política de contenido, Meta modificó la política de personas y organizaciones peligrosas para permitir contenido que incluya el término "shaheed" en todos los idiomas que lo usan, excepto cuando el contenido va acompañado de señales de violencia o, de otro modo, infringe nuestras políticas (por ejemplo, porque glorifica a personas peligrosas designadas).

En 2024, el Consejo asesor de contenido también probó acortar los plazos para los casos urgentes. Por ejemplo, después de las elecciones presidenciales de Venezuela en julio, cuando se desató la violencia, remitimos dos elementos de contenido respecto de los "colectivos" para que se sometan a una revisión inmediata. "Colectivos" es un término general que describe a pandillas armadas irregulares o grupos de estilo paramilitar que mantienen un estrecho vínculo con el Gobierno. Se aceleró el plazo a 14 días para tomar una decisión en estos casos.

Asimismo, nos asociamos con el Consejo asesor de contenido para hacer partícipes a autoridades reguladoras y organizaciones de la sociedad civil, incluido en la región de África, Latinoamérica, Oriente Medio y Turquía, para generar consciencia acerca de la orden y el proceso de selección de casos del Consejo asesor de contenido.

# Anexo

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

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 (ahora director general 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 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.

Nuestra Política corporativa de derechos humanos nos insta a presentar ante la junta directiva informes periódicos sobre asuntos clave relacionados con los derechos humanos. En 2024, el director de derechos humanos dio instrucciones al Comité de Auditoría y Supervisión de Riesgos del Consejo.

En 2024, Meta lanzó el sector de riesgos para los derechos humanos del programa de gestión de riesgos de terceros de la empresa. Este control demuestra nuestro compromiso con mejorar continuamente nuestra gestión de los riesgos que ponen en peligro los derechos humanos y esforzarnos para forjar colaboraciones con terceros que honren las responsabilidades para con los derechos humanos y los respeten.

## 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 2024. 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

Informe de prácticas comerciales responsables de 2025

Informe de sustentabilidad de 2025

Informe sobre derechos humanos de 2023, 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 2024

Informe sobre minerales en zonas de conflicto de 2024

Informes de transparencia de Meta

Informes regulatorios y otros informes de transparencia

Evaluaciones sobre el impacto en los derechos humanos publicadas anteriormente: Cifrado de extremo a extremo, Filipinas, Myanmar, Indonesia, Camboya, India, Sri Lanka e Israel y Palestina


Meta