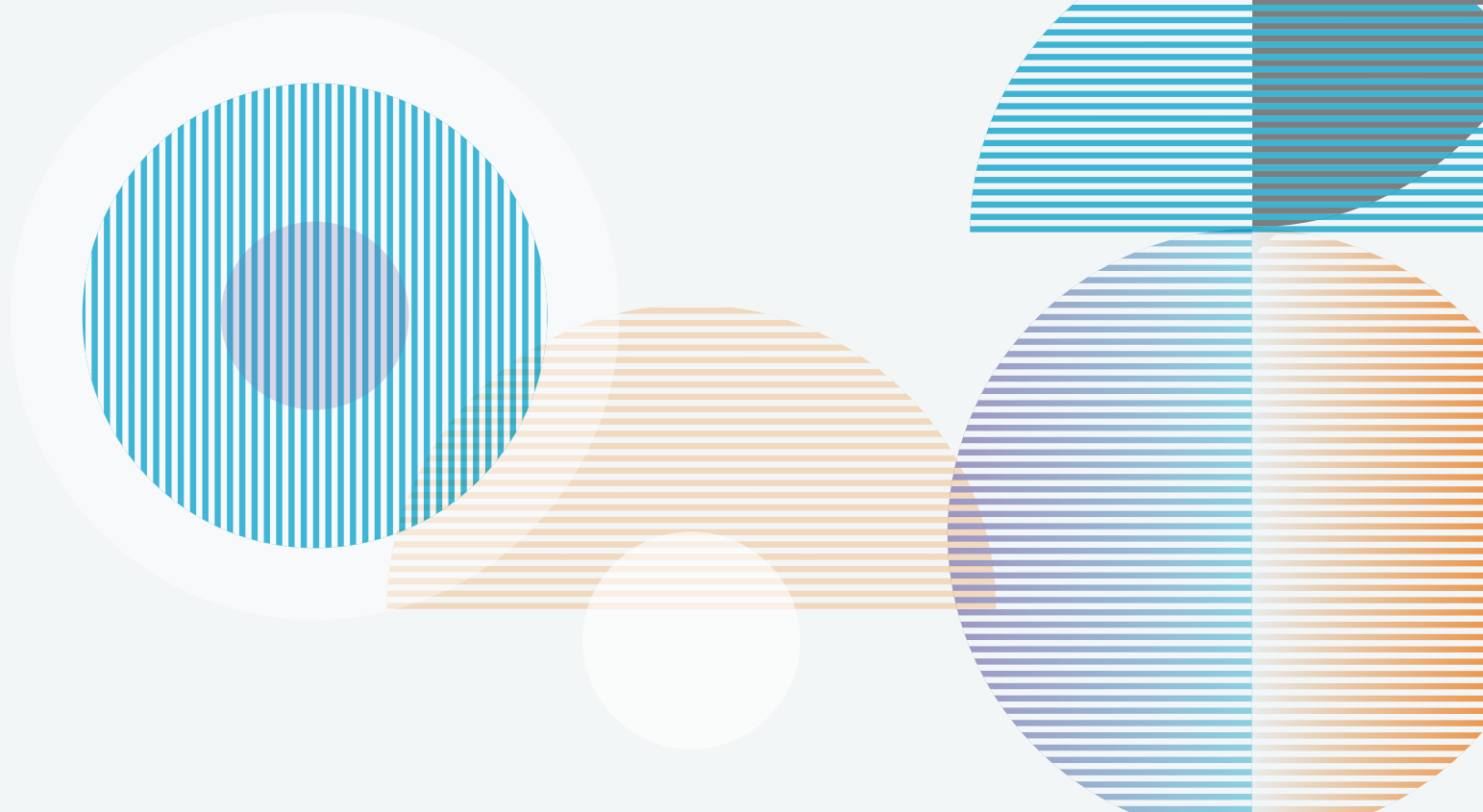




Oversight Board

**INFORME
ANUAL DE**

2022



Índice general

Prefacio de los copresidentes	3
Prefacio del presidente del fideicomiso	7
Resumen ejecutivo	8
Conozca al Consejo	10
Presentación de nuestras siete prioridades estratégicas	12
Cómo considera el Consejo las apelaciones de los usuarios	16
Cronología de eventos clave en 2022	18

Recomendaciones e impacto

Información general	21
Recomendaciones e impacto en 2022	22
De compromisos a la acción: resultados para los usuarios	23
Cronología del impacto: informar a los usuarios qué hicieron mal	28
Implementación de nuestras recomendaciones por parte de Meta	30

Selección de casos

Información general	33
Casos presentados al Consejo	34
Casos que consideró el Comité de selección de casos	37

Decisiones de casos y opiniones consultivas sobre políticas

Información general	39
Decisiones y opiniones consultivas sobre políticas emitidas en 2022	41
Resúmenes de decisiones y opiniones consultivas sobre políticas	42
Aplicación de normas internacionales sobre derechos humanos a la moderación de contenido: artículo 19	54
Normas internacionales sobre derechos humanos en el proceso de toma de decisiones del Consejo	55

Participación y comentarios públicos

Información general	59
Cronología de las actividades de participación en 2022	62

Próximos pasos

Evolucionar nuestro trabajo con Meta	63
Compartir los beneficios de la supervisión independiente	65
Ayudar a las empresas a adaptarse a la normativa emergente	67

Prefacio de los copresidentes




Evelyn Aswad, Catalina Botero-Marino, Michael McConnell, Helle Thorning-Schmidt: COPRESIDENTES DEL CONSEJO ASESOR DE CONTENIDO

En 2022, muchas de nuestras recomendaciones para Meta se volvieron una realidad, lo cual mejoró el trato de la empresa hacia las personas y las comunidades de todo el mundo. Nuestro trabajo produjo que Meta revisara sus políticas de moderación de contenido, estableciera sus reglas con más claridad y las aplicara con más coherencia. Meta ahora indica a más usuarios qué área específica de la política infringieron cuando se eliminan sus publicaciones y alinea mejor su moderación de contenido con los principios de los derechos humanos.

En respuesta a nuestras recomendaciones, Meta presentó el Protocolo sobre política de crisis para que sus respuestas a las situaciones de crisis sean más coherentes, lanzó una revisión de su política sobre personas y organizaciones peligrosas, y creó una nueva norma comunitaria sobre información errónea. En respuesta a una recomendación relacionada con nuestra decisión sobre “síntomas de cáncer de mama y desnudos”, Meta también mejoró sus técnicas para identificar el contexto del cáncer de mama en el contenido de Instagram, lo cual contribuyó a que miles de publicaciones adicionales se enviaran para su revisión manual, las cuales anteriormente se habrían eliminado automáticamente.

En 2022, hicimos más de la mitad de las 91 recomendaciones de políticas como parte de nuestras primeras opiniones consultivas sobre políticas, incluida una sobre cómo Meta trata a sus usuarios más importantes en su programa de verificación. Como respuesta, Meta se comprometió a aumentar la protección a aquellos que corren un riesgo especial de aplicación excesiva, incluidos los periodistas y los defensores de los derechos humanos.

También protegimos la voz de los usuarios, especialmente durante las transformaciones y crisis políticas y sociales. Por ejemplo, a principios de 2023, como parte de una decisión específica del Consejo, instamos a Meta a proteger mejor el discurso político en Irán, donde protestas generalizadas e históricas se han reprimido con violencia. Como respuesta, Meta permitió que el término “Marg bar Khamenei” (que literalmente se traduce como “muerte al [líder supremo de Irán] Khamenei”) se compartiera en el contexto de las protestas en curso en Irán. Meta también cambió su sistema de faltas y sanciones para que sea más justo con los usuarios.



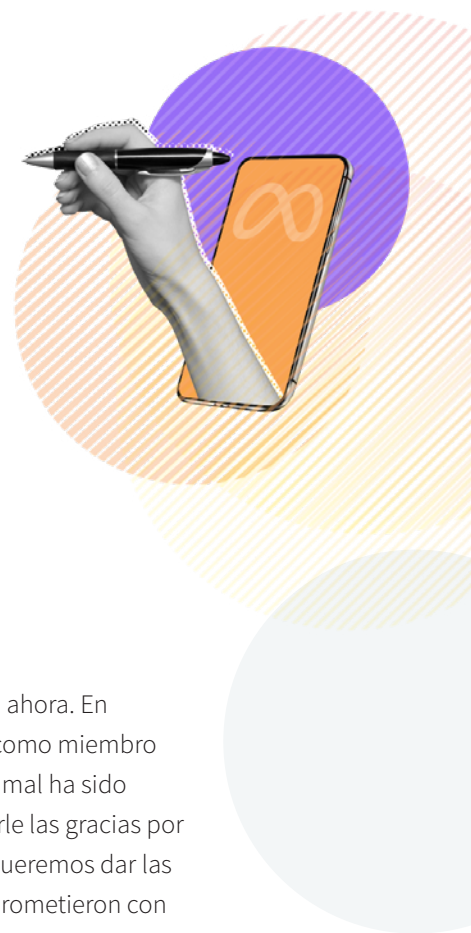
Cada decisión y opinión consultiva sobre políticas aportó más transparencia a los procesos de moderación de contenido, con frecuencia opacos, revelando la cantidad de excepciones de interés periodístico que la empresa aplica en la administración de sus reglas, entre otras medidas. Nuestras recomendaciones de políticas generan el debate público sobre cómo las plataformas digitales pueden abordar algunos de los retos más complejos de la moderación de contenido.

Para aumentar nuestro impacto, adoptamos siete áreas prioritarias en las que queremos trabajar con las partes interesadas para mejorar las experiencias de las personas en línea. Estas áreas son elecciones y espacio cívico, situaciones de crisis y conflicto, género, lenguaje que incita al odio contra grupos marginados, uso gubernamental de las plataformas de Meta, trato justo hacia los usuarios, y aplicación automatizada de políticas y selección del contenido. Nos preparamos también para asumir un número de casos más grande y tomar decisiones más rápido en 2023.

En 2022, también notamos que obtenía cada vez más reconocimiento la idea de que las empresas no deben tomar solas las decisiones sobre la moderación de contenido. Desde el principio, el Consejo se diseñó para probar un enfoque independiente de la moderación de contenido, que, de tener éxito, también se podría aplicar a otras empresas. La supervisión independiente consiste en que las empresas se abran e inviten a personas ajenas a cuestionar su funcionamiento. En los últimos tres años, hemos adquirido una gran experiencia en materia de supervisión independiente que puede ayudar a las empresas a tomar decisiones más sólidas basadas en el respeto de la libertad de expresión y otros derechos humanos. Como la nueva normativa trae consigo nuevos requisitos, también hay áreas específicas, como la transparencia y las notificaciones de usuario, en las que consideramos que podemos aportar parte de la solución.

Nos gustaría aprovechar esta oportunidad para dar las gracias a los fiduciarios del Consejo asesor de contenido, al personal de la Administración y a nuestros compañeros miembros del Consejo por su experiencia y apoyo en nuestro camino hasta ahora. En particular, nos gustaría reconocer la contribución de Jamal Greene, que dejó su puesto como miembro del Consejo asesor de contenido y copresidente en diciembre de 2022. El liderazgo de Jamal ha sido fundamental para nuestro éxito a la hora de hacer responsable a Meta, y nos gustaría darle las gracias por todo lo que ha hecho para establecer el Consejo y avanzar en nuestra misión. También queremos dar las gracias a las numerosas partes interesadas que enviaron comentarios públicos, se comprometieron con nuestro trabajo y contribuyeron a hacer posibles nuestros logros hasta la fecha.

Dado el inexplorado camino que estamos recorriendo, el Consejo asesor de contenido sigue adaptándose y encontrando nuevas formas de cumplir nuestra misión. Aunque hemos avanzado mucho hasta ahora, no nos hacemos ilusiones sobre la magnitud del reto que tenemos por delante. Juntos podemos ayudar a superar los inconvenientes de las redes sociales y a las personas a conectarse con confianza.





Prefacio del presidente del fideicomiso



Stephen Neal
PRESIDENTE DEL
FIDEICOMISO DEL
CONSEJO ASESOR
DE CONTENIDO

2022 fue mi primer año completo como presidente del fideicomiso del Consejo asesor de contenido, y quedé muy impresionado por el trabajo del Consejo. Los miembros del Consejo siguieron deliberando sobre los casos más difíciles y significativos, así como sobre decisiones determinantes en una serie de asuntos. Estos incluyeron la publicación de la primera opinión consultiva sobre políticas del Consejo, que examinó el uso compartido de la información de residencia privada, y la decisión sobre el “poema ruso” relacionada con la invasión a Ucrania. Como fiduciarios, ayudamos a nombrar a tres nuevos miembros del Consejo de Egipto, México y Estados Unidos.

Otro aspecto crucial de nuestro rol es supervisar a la Administración del Consejo asesor de contenido, que cuenta con personal a tiempo completo dedicado a ayudar a los miembros del Consejo con su trabajo. En 2022, la Administración completó la contratación de todos los equipos y ahora cuenta con, aproximadamente, 80 personas ubicadas en Londres, Washington D. C. y San Francisco. Hemos atraído a nuevos colegas excelentes, y muchos de ellos tienen una experiencia única en libertad de expresión y derechos humanos. Al igual que las redes sociales, la Administración es mundial, y sus empleados hablan 40 idiomas.

En julio de 2022, anunciamos un compromiso adicional de USD 150.000.000 por parte de Meta, además de los USD 130.000.000 anunciados en 2019 cuando se estableció el fideicomiso. Al crear un compromiso financiero continuo, Meta emitió un voto de confianza en el trabajo del Consejo y sus esfuerzos para aplicar las normas de contenido de Facebook e Instagram de una manera que proteja la libertad de expresión y las normas pertinentes sobre derechos humanos.

En 2023, seguiremos supervisando el funcionamiento del Consejo y salvaguardando su independencia, dos aspectos fundamentales para su éxito. Mediante sus decisiones de casos y opiniones consultivas sobre políticas, el Consejo continuará mejorando los productos y las políticas de Meta, lo cual da lugar a una mejor experiencia para quienes usan Facebook e Instagram. Mediante la colaboración con grupos de la sociedad civil, reguladores y otras plataformas, el Consejo aspira a reforzar su legitimidad. Los empleados de Meta son otra agrupación crucial con una gran influencia en el futuro de la empresa, que el Consejo intentará aprovechar como defensores de su labor.

Meta merece crédito por su visión al establecer el Consejo como una nueva forma de gobernanza de las redes sociales y por su compromiso continuo con este esfuerzo. Como fiduciarios, apoyaremos el éxito continuo del Consejo con Meta, y ayudaremos al Consejo a compartir su enfoque con otras empresas y socios en el sector.

“**Al crear este compromiso financiero continuo, Meta emitió un voto de confianza en el trabajo del Consejo”.**

Stephen Neal
PRESIDENTE DEL FIDEICOMISO DEL CONSEJO ASESOR DE CONTENIDO

Resumen ejecutivo

En 2022, el Consejo asesor de contenido hizo 91 recomendaciones a Meta.

En respuesta a nuestras recomendaciones hasta el momento, Meta:



Comenzó a indicar a los usuarios qué política específica infringía su contenido al momento de eliminarlo.



Empezó sistemáticamente a medir la transparencia de su mensajería de aplicación a los usuarios.



Mejóro la manera de identificar el contexto del cáncer de mama en el contenido de Instagram, lo cual contribuyó a que miles de publicaciones adicionales se enviaran para su revisión manual, las cuales anteriormente se habrían eliminado automáticamente.



Finalizó la implementación mundial de la nueva mensajería que indica a los usuarios si la revisión manual o automatizada ocasionó que su contenido se eliminara.



Creó una nueva sección en las normas comunitarias sobre información errónea.



Presentó el nuevo Protocolo sobre política de crisis.

En 2022, el Consejo asesor de contenido:

Recibió cerca de **1,3m** de casos de usuarios de todo el mundo, casi un cuarto más que en 2021.

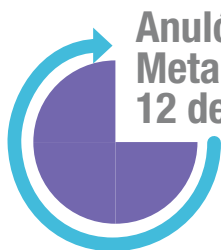
Emitió las primeras opiniones consultivas sobre políticas,



Las cuales cubren el uso compartido de información de residencia privada



El programa de verificación de Meta.



Anuló las decisiones de Meta en tres cuartos de las 12 decisiones de casos

Es decir que anuló sus decisiones de moderación de contenido 9 veces y las confirmó 3 veces.

Publicó **12** decisiones de casos

sobre temas que van desde la invasión a Ucrania por parte de Rusia hasta la influencia de las autoridades policiales en la eliminación de contenido.

Amplió nuestro ámbito para poder agregar pantallas de advertencia a contenido elegible.



Hizo que Meta revirtiera su decisión original en

32 casos considerados para su selección

en los que su decisión original sobre una publicación era incorrecta

En 2023, haremos lo siguiente:

- ✓ Publicaremos nuestras primeras decisiones sumarias sobre los casos en los que Meta revirtió su decisión original sobre un contenido.
- ✓ Emitiremos nuestras primeras decisiones expeditas en las que publicamos una decisión sobre un caso en días.
- ✓ Alcanzaremos nuestro objetivo actualizado del Consejo en pleno para lograr la máxima eficacia.
- ✓ Reforzaremos el compromiso en torno a nuestras siete prioridades estratégicas.
- ✓ Proseguiremos los planes a largo plazo para ampliar el ámbito.
- ✓ Supervisaremos la forma en que Meta implementa nuestras recomendaciones y pediremos a la empresa que proporcione pruebas de su implementación e impacto.

Creemos en el valor de la supervisión independiente y exploraremos la posibilidad de nuevas asociaciones con las empresas; también analizaremos la forma en que nuestro trabajo puede complementar mejor la normativa emergente.

Conozca al Consejo

Miembros del Consejo asesor de contenido



Afia Asantewaa Asare-Kyei

Director de Responsabilidad y Justicia,
Open Society Foundations-Africa



Tawakkol Karman

Premio Nobel de la Paz



Evelyn Aswad

Profesora y catedrática, Facultad
de Derecho de la Universidad de
Oklahoma



Sudhir Krishnaswamy

Vicerrector y profesor de Derecho,
Facultad Nacional de Derecho de la
Universidad de la India



Endy Bayuni

Editor sénior y miembro del
directorio, The Jakarta Post



Ronaldo Lemos

Profesor, Facultad de Derecho de
la Universidad del Estado de Río de
Janeiro



Catalina Botero-Marino

Directora de la cátedra de libertad
de expresión de la UNESCO en la
Universidad de los Andes



Khaled Mansour

Escritor



Paolo Carozza

Profesor, Universidad de Notre Dame



Michael McConnell

Profesor y director del Centro de
Derecho Constitucional, Facultad
de Derecho de Stanford



Katherine Chen

Profesora, Universidad Nacional
de Chengchi



Suzanne Nossel

Directora ejecutiva de PEN America



Nighat Dad

Fundadora, Digital Rights Foundation



Julie Owono

Directora ejecutiva, Internet Sans
Frontières



Emi Palmor

Abogada y profesora universitaria,
Centro Interdisciplinario Herzliya,
Israel



Pamela San Martín

Ex consejera electoral del Instituto
Nacional Electoral (INE) de México



Alan Rusbridger

Director, Lady Margaret Hall, Oxford



Nicolas Suzor

Profesor, Facultad de Derecho de
la Universidad de Tecnología de
Queensland



András Sajó

Profesor universitario, Universidad
Centroeuropa



Helle Thorning-Schmidt

Ex primera ministra, Dinamarca



John Samples

Vicepresidente, Instituto Catón



Kenji Yoshino

Profesor “Chief Justice Earl Warren”
de Derecho Constitucional y director
del Meltzer Center for Diversity,
Inclusion, and Belonging

Fiduciarios del Consejo asesor de contenido



Kristina Arriaga

Fiduciaria



Kate O'Regan

Fiduciaria



Cherine Chalaby

Fiduciario



Robert Post

Fiduciario



Stephen Neal

Presidente del fideicomiso del
Consejo asesor de contenido



Marie Wieck

Fiduciaria

Administración del Consejo asesor de contenido



Thomas Hughes

Director

Presentación de nuestras siete prioridades estratégicas

En octubre de 2022, anunciamos siete prioridades estratégicas en función de un análisis exhaustivo y profundo de las cuestiones planteadas por las apelaciones de los usuarios al Consejo. Dado que estas prioridades ahora guían los casos que seleccionamos, animamos a los usuarios a tomarlas en cuenta al enviar apelaciones.



1. Elecciones y espacio cívico

Las empresas de redes sociales se enfrentan a retos cuando deben aplicar de manera coherente sus políticas a la expresión política en muchas partes del mundo, incluso durante elecciones y protestas a gran escala. Resaltamos la importancia de proteger la expresión política en nuestra decisión de “manifestaciones en Rusia a favor de Navalny”, mientras que nuestra decisión sobre la “mención de los talibanes en los informes de noticias” abordó cuestiones de la libertad de los medios de comunicación. Como Consejo, nos gustaría explorar las responsabilidades de Meta en las elecciones, protestas y otros momentos clave para la participación ciudadana.



2. Situaciones de crisis y conflicto

En tiempos de crisis, como conflictos armados, ataques terroristas y emergencias sanitarias, las redes sociales pueden ayudar a las personas a intercambiar información fundamental, debatir sobre importantes cuestiones públicas y mantenerse a salvo. Sin embargo, también pueden crear un entorno donde se difunda información errónea y odio. En nuestras decisiones de “presuntos delitos en Raya Kobo” y del “Consejo de comunicaciones de Tigray”, se examinaron las publicaciones relacionadas con el conflicto en Etiopía, mientras que nuestra decisión sobre el expresidente Trump produjo que Meta adoptara el Protocolo sobre política de crisis. Como Consejo, nos gustaría explorar el rol de Meta en la protección de la libertad de expresión en tales circunstancias, así como su preparación ante posibles daños a los que puedan contribuir sus productos durante conflictos armados, disturbios civiles y otras emergencias.



3. Género

Las mujeres y las personas trans y no binarias se enfrentan a obstáculos que les impiden ejercer su derecho a la libertad de expresión en las redes sociales. En nuestra decisión de “síntomas de cáncer de mama y desnudos”, por ejemplo, los sistemas automatizados de Meta no pudieron aplicar las excepciones para la concientización sobre el cáncer de mama, lo cual produjo que se eliminara de Instagram información importante sobre la salud. En nuestra decisión sobre “identidad de género y desnudos”, que se publicó a principios de 2023, también se concluyó que las políticas de Meta sobre desnudos de adultos se traducen en mayores barreras para la expresión de mujeres y personas trans y no binarias en Facebook e Instagram. Como Consejo, nos gustaría

explorar los obstáculos relacionados con el género a los que se enfrentan las mujeres y las personas de la comunidad LGBTQIA+ al momento de ejercer su derecho a la libertad de expresión, incluida la violencia de género y el acoso, y los efectos de las distinciones de género en las políticas de contenido.



4. Lenguaje que incita al odio contra grupos marginados

El lenguaje que incita al odio crea un entorno de discriminación y hostilidad hacia grupos marginados. Usualmente tiene que ver con el contexto, está codificado y es dañino, y se deriva de efectos que se acumulan poco a poco con el paso del tiempo. En nuestra decisión de la “representación de Zwarte Piet”, se concluyó que permitir que imágenes de caras pintadas de negro se acumularan en Internet crearía un entorno discriminatorio para las personas negras. Por otro lado, en nuestras decisiones sobre el “cinturón de Wampum” y la “reapropiación de términos árabes”, se examinó el “contradiscurso”, que hace referencia al lenguaje que incita al odio para resistirse a la discriminación. Como Consejo, nos gustaría explorar cómo Meta debe proteger a los miembros de grupos marginados, además de garantizar que su aplicación no se oriente de manera equivocada hacia quienes se enfrentan al odio. Al mismo tiempo, sabemos que las restricciones al lenguaje que incita al odio no se deben aplicar en exceso ni utilizar para limitar el ejercicio legítimo de la libertad de expresión, incluida la expresión de puntos de vista poco populares o controversiales.



5. Uso gubernamental de las plataformas de Meta

Los Gobiernos usan Facebook e Instagram para transmitir sus políticas y hacer solicitudes a Meta para que elimine contenido. En respuesta a nuestra decisión del “aislamiento de Öcalan”, Meta aceptó proporcionar información sobre el contenido eliminado por infringir las normas comunitarias después de una denuncia de un Gobierno. En nuestra decisión sobre “música drill británica”, también se hicieron propuestas sobre cómo debe responder Meta a las solicitudes de autoridades policiales nacionales. Como Consejo, nos gustaría explorar cómo los agentes estatales usan las plataformas de Meta, cómo estos podrían influir en las prácticas y las políticas de moderación de contenido (a veces, de maneras que no son transparentes), y las implicaciones de la participación del estado en la moderación de contenido.



6. Trato justo hacia los usuarios

Cuando el contenido de las personas se elimina de Facebook e Instagram, no siempre se les dice qué regla infringieron. En otras instancias, no se trata de igual manera a los usuarios o no se les dan garantías procesales adecuadas y acceso a soluciones para los errores cometidos. Como Consejo, nos gustaría explorar cómo Meta puede tratar mejor a sus usuarios, proporcionando notificaciones de usuario más específicas, garantizando que las personas siempre puedan apelar la decisión de Meta ante la empresa y siendo más transparente en áreas como las “faltas” y la verificación.



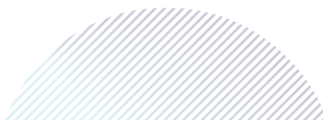
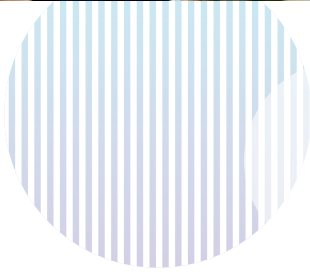
7. Aplicación automatizada de políticas y selección del contenido

Si bien los algoritmos son cruciales para la moderación de contenido a escala, hay una falta de transparencia y comprensión en torno a cómo funcionan los sistemas automatizados de Meta y cómo afectan el contenido que ven los usuarios. En nuestra decisión de la “caricatura sobre la policía colombiana”, se demostró cómo la automatización puede aumentar el impacto de decisiones incorrectas de moderación de contenido. En respuesta a nuestra decisión de “síntomas de cáncer de mama y desnudos”, Meta implementó una nueva mensajería en todo el mundo que indica a los usuarios si la revisión manual o automatizada ocasionó que su contenido se eliminara. Como Consejo, nos gustaría explorar cómo se debe designar y revisar la aplicación automatizada, la precisión y las limitaciones de los sistemas automatizados, y la importancia de una mayor transparencia en esta área.

||| TRABAJO CON LAS PARTES INTERESADAS PARA AUMENTAR NUESTRO IMPACTO

Como Consejo, nuestros logros hasta el momento han sido posibles gracias a escuchar a investigadores, grupos de la sociedad civil y otras personas que han trabajado durante muchos años en los problemas a los que nos enfrentamos, y gracias a colaborar con estos. A fin de encontrar soluciones prácticas para nuestras prioridades estratégicas, y los enormes retos que plantean, es esencial la experiencia en el tema en cuestión y el conocimiento local de estas partes interesadas.

En el caso de todas las prioridades estratégicas, continuaremos trabajando con un gran número de partes interesadas que reflejan la diversidad de las personas que usan las plataformas de Meta. Esto nos ayudará a entender las políticas y las prácticas de aplicación que Meta necesita mejorar con mayor urgencia, y qué tipo de casos podrían brindar la oportunidad de abordarlas. Queremos asociarnos con organizaciones en todo el mundo para lograrlo, mediante nuestros procesos de comentarios públicos, mesas redondas y conversaciones individuales. Para hablar sobre cómo puede participar su organización, póngase en contacto con engagement@osbadmin.com.



Cómo considera el Consejo las apelaciones de los usuarios

En este gráfico se indica el proceso de apelación tal como se aplicó a las decisiones sobre las apelaciones de los usuarios en 2022.





DECISIÓN

El panel toma una decisión sobre si se debe permitir el contenido, y confirmar o anular la decisión de Meta.



PUBLICACIÓN

Nuestra decisión se publica en el sitio web del Consejo asesor de contenido. Meta tiene que implementar esta decisión en un plazo de 7 días a partir de la publicación y responder a las recomendaciones en un plazo de 60 días.



APROBACIÓN

Un borrador de la decisión se distribuye a todos los miembros del Consejo para su revisión. La mayoría debe firmarlo para que se publique una decisión.

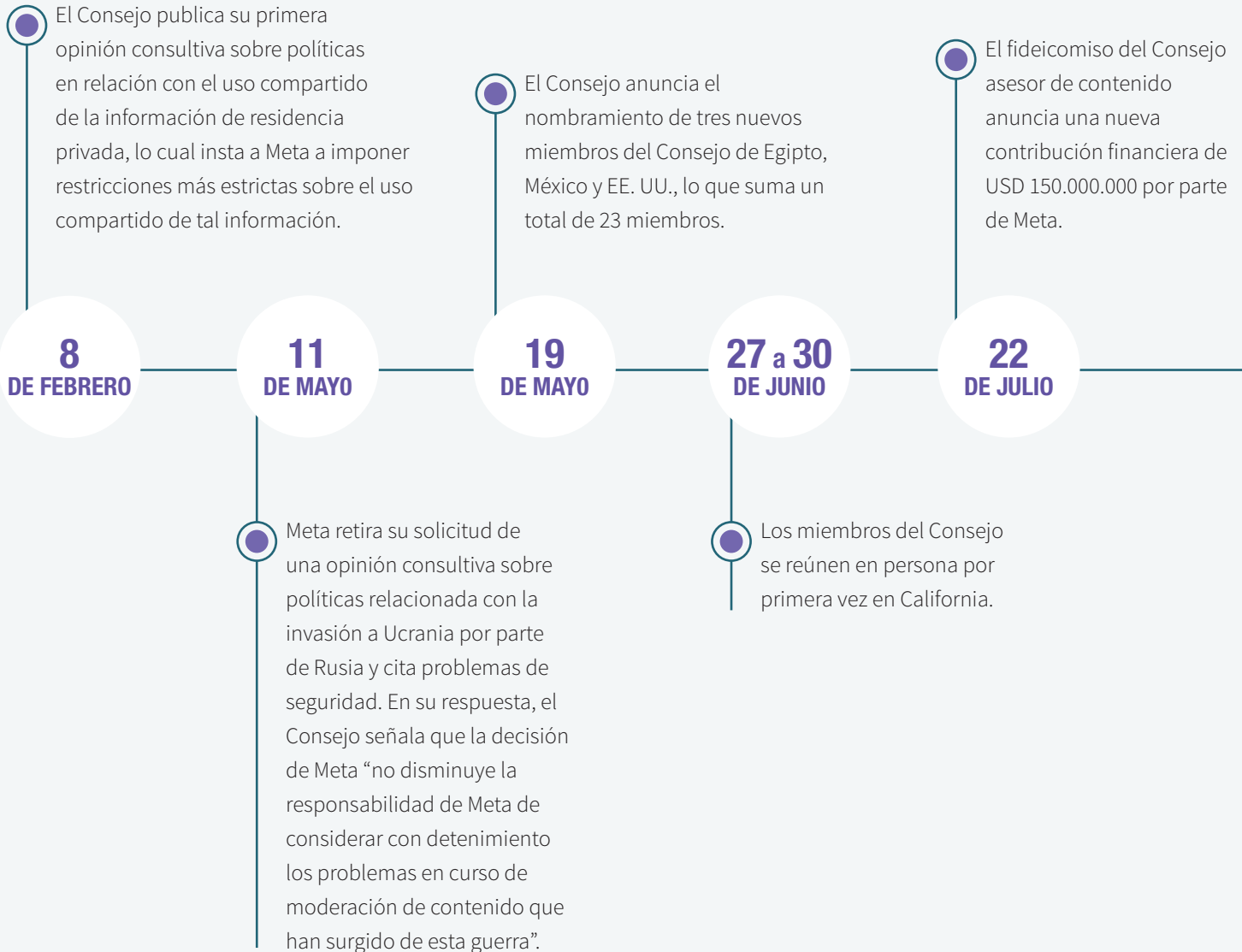


IMPLEMENTACIÓN

El Consejo supervisa cómo Meta implementa las recomendaciones y brinda actualizaciones en informes de transparencia trimestrales.



Cronología de eventos clave en 2022



**26
DE JULIO**

El Consejo acepta la solicitud de Meta para obtener una opinión consultiva sobre políticas respecto de la eliminación de información errónea sobre el COVID-19.

**20
DE OCTUBRE**

El Consejo anuncia siete prioridades estratégicas centradas en áreas donde puede tener el mayor impacto en las experiencias de las personas en Facebook e Instagram.

**20
DE OCTUBRE**

El Consejo puede aplicar pantallas de advertencia que marcan publicaciones como “perturbadoras” o “sensibles” al restaurar contenido elegible o al dejarlo.

**22 DE
NOVIEMBRE**

El Consejo publica su decisión sobre “música drill británica”; es la primera vez que ha examinado una publicación eliminada después de una solicitud de una autoridad policial nacional.

**6 DE
DICIEMBRE**

El Consejo publica una opinión consultiva sobre políticas relacionada con el programa de verificación de Meta. Encuentra que la verificación es deficiente en áreas clave y hace 32 propuestas a Meta.

Recomendaciones e impacto

Hicimos **91** recomendaciones a Meta en 2022.

En respuesta a nuestras recomendaciones hasta el momento, Meta:



Lanzó notificaciones nuevas a nivel mundial que indican a los usuarios la política específica que infringieron respecto de sus políticas de lenguaje que incita al odio, personas y organizaciones peligrosas, y bullying y acoso.



Comenzó sistemáticamente a medir la transparencia de su mensajería de aplicación a los usuarios.



Mejóro la forma de identificar el contexto del cáncer de mama en el contenido de Instagram, lo cual contribuyó a que miles de publicaciones adicionales se enviaran para su revisión manual, las cuales anteriormente se habrían eliminado automáticamente.



Finalizó la implementación mundial de la mensajería para los usuarios que indica a las personas si la revisión manual o automatizada ocasionó que su contenido se eliminara.



Lanzó notificaciones nuevas que indican a los usuarios si se ha restringido su acceso al contenido debido a una ley local, luego de una solicitud del Gobierno.



Creó una nueva sección en las normas comunitarias sobre información errónea.



Lanzó el Protocolo sobre política de crisis.



Empezó una revisión profunda de su política de personas y organizaciones peligrosas para priorizar las designaciones en función del riesgo.



Información general

En nuestras decisiones de casos y opiniones consultivas sobre políticas, ofrecemos recomendaciones específicas sobre cómo Meta puede mejorar las políticas que aplica al contenido de miles de millones de usuarios. Aunque nuestras recomendaciones no son vinculantes, Meta debe responder a ellas públicamente en un plazo de 60 días. Meta ha reconocido públicamente cómo nuestras recomendaciones están cambiando su comportamiento. En agosto de 2022, la empresa afirmó que el Consejo “continúa presionándonos para ser más cuidadosos sobre el impacto de nuestra moderación de contenido mundial y más equitativos en nuestra aplicación de políticas y uso de recursos. Sobre todo, también nos presionan para ser más transparentes, ya que las voces externas pueden sernos de ayuda para responsabilizarnos de nuestras promesas”.

Como Consejo, responsabilizamos a Meta publicando informes de transparencia cada trimestre. Estos aplican un enfoque riguroso, independiente y basado en datos para evaluar el progreso de Meta en la implementación de nuestras recomendaciones con el transcurso del tiempo. Al hacer públicamente estas recomendaciones, y supervisar públicamente las respuestas y la implementación de Meta, hemos abierto un espacio para el diálogo transparente con la empresa que no existía anteriormente. Este tipo de transparencia ayuda a crear legitimidad y confianza con los usuarios y la sociedad civil.

El rol de los grupos de la sociedad civil en el desarrollo de nuestras recomendaciones tampoco se puede exagerar. En muchos casos, estas organizaciones envían ideas específicas para las recomendaciones como parte de nuestro proceso de comentarios públicos. En otros casos, nuestras propuestas repercuten o se basan en llamamientos que estos grupos han hecho durante muchos años, lo cual obliga a Meta a tener en cuenta estos antiguos llamamientos a la acción y responder a estos públicamente. Aunque mencionamos explícitamente estas influencias en nuestros textos de decisiones, nos gustaría reiterar nuestra gratitud hacia estas organizaciones por compartir sus ideas y su experiencia.



[Las recomendaciones del Consejo] también nos presionan para ser más transparentes, ya que las voces externas pueden sernos de ayuda para responsabilizarnos de nuestras promesas”.

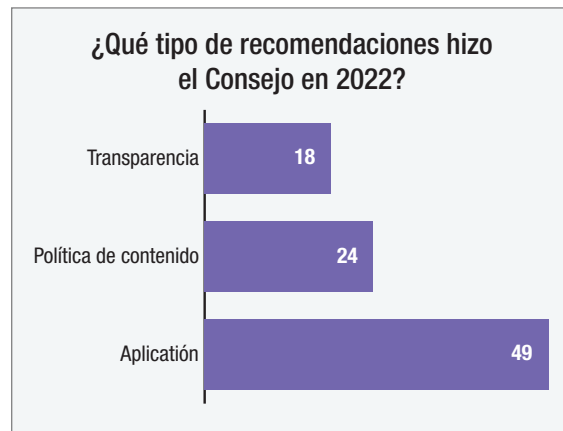
Actualización trimestral del segundo trimestre de 2022 de Meta sobre el Consejo asesor de contenido

Recomendaciones e impacto en 2022

El Consejo hizo 91 recomendaciones a Meta en 2022, cantidad ligeramente superior a las 86 propuestas que hicimos a la empresa en 2021. Para principios de abril de 2023, habíamos hecho otras 14 recomendaciones, lo que da un total de 191 recomendaciones a Meta.

En total, 41 recomendaciones se clasificaron en las categorías de «implementación demostrada» o «implementación parcial demostrada», y evaluamos otras 84 recomendaciones como «progreso informado».

En 2022, fue particularmente alentador ver que, al implementar nuestras recomendaciones, Meta hizo varios cambios que tuvieron un impacto sistémico en el enfoque de la empresa. Entre estos, se incluye la implementación de notificaciones más específicas para los usuarios, que pedimos desde enero de 2021. Como resultado de recomendaciones reiteradas del Consejo, Meta también actualizó la forma de medir la especificidad y transparencia de la mensajería que proporciona a los usuarios cuando toma medidas de aplicación contra el contenido por infringir sus políticas. Este cambio sistémico forma parte de los esfuerzos generales de la empresa para ser más específica con los usuarios. Meta también está llevando a cabo una revisión profunda de la definición de “elogio” en relación con su marco de elogio, apoyo sustantivo y representación (PSR) en la política de personas y organizaciones peligrosas. Meta usa este marco para evaluar cómo se representan de manera positiva las personas y organizaciones peligrosas en el contenido de los usuarios.



LECCIONES APRENDIDAS

Un área en la que la cooperación con Meta se podría haber mejorado en 2022 fue el acceso a los datos. A finales de 2021 y principios de 2022, dedicamos ocho meses a intentar obtener acceso a la herramienta CrowdTangle de Meta para que nos diera más información al seleccionar los casos y evaluar el impacto de la recomendación. Después de encontrar varios obstáculos, presentamos este problema al liderazgo de Meta a principios de 2022 y finalmente se nos concedió acceso. Meta tomó varias medidas positivas sobre compartir datos más adelante en 2022, por ejemplo, compartir investigaciones en curso sobre las apelaciones de los usuarios y contratar a un científico de datos para validar la implementación de nuestras recomendaciones. Esperamos continuar nuestra asociación con los equipos de ciencia de datos de Meta para obtener datos significativos que demuestren pruebas de la implementación y del impacto.

PUNTO DESTACADO DE LA RECOMENDACIÓN

Un capítulo nuevo en la comprensión de nuestro impacto

En nuestra decisión de “síntomas de cáncer de mama y desnudos”, publicada en enero de 2021, recomendamos que Meta **“mejore la detección automatizada de imágenes con superposiciones de texto para garantizar que las publicaciones que busquen concientizar acerca de los síntomas del cáncer de mama no se marquen de forma errónea para su revisión”**. En respuesta, el equipo de implementación de Meta mejoró las técnicas de Instagram para identificar contenido en el contexto del cáncer de mama a través de texto y las implementó en julio de 2021. Estas mejoras se han aplicado desde entonces y, en los 30 días comprendidos entre el 26 de febrero y el 27 de marzo de 2023, contribuyeron a que se enviaran 2.500 contenidos para su revisión manual, los cuales anteriormente se habrían eliminado.

Aunque es difícil contextualizar 2.500 contenidos sin un denominador, podemos ver que Meta se ciñó a nuestro marco al implementar la recomendación, y ello redujo de manera satisfactoria la aplicación excesiva en la plataforma. Esto es una victoria para la gobernanza independiente y el comienzo de un capítulo nuevo para la colaboración entre Meta y el Consejo sobre la comprensión de nuestro impacto en los sistemas de Meta.

De compromisos a la acción: resultados para los usuarios

Dada la ambición de nuestras recomendaciones y los cambios técnicos que a menudo requieren, comprendemos que tardan en implementarse. En 2022, notamos progreso en muchas de las recomendaciones que hicimos en 2021, así como compromisos nuevos en respuesta a las propuestas más recientes. Fue alentador ver que, por primera vez, Meta aprobó cambios sistémicos en sus reglas y cómo se aplican, incluidas las de notificaciones de usuario y sus reglas de organizaciones peligrosas. Los ejemplos siguientes demuestran el impacto de nuestras recomendaciones sobre cómo trata la empresa a los usuarios y a las comunidades en todo el mundo.



||| CAMBIOS SISTÉMICOS EN LAS REGLAS Y LA APLICACIÓN DE META

- ◉ Como Consejo, la recomendación que hemos hecho más a menudo es que Meta **indique a las personas qué hicieron mal** cuando se elimina su contenido. Desde la primera vez que hicimos esta recomendación en enero de 2021, Meta ha avanzado poco a poco hacia este objetivo. En respuesta a esta recomendación, Meta presentó una nueva mensajería en todo el mundo que indica a los usuarios la política específica que infringieron respecto de sus políticas de lenguaje que incita al odio, personas y organizaciones peligrosas, y bullying y acoso. En respuesta a nuestra recomendación, Meta ahora también mide sistémicamente el nivel de detalle de las comunicaciones con los usuarios en todas las eliminaciones de contenido.
- ◉ En respuesta a una recomendación de nuestra decisión sobre el expresidente Trump, Meta ha hecho **cambios sistémicos en su respuesta a las crisis y los conflictos**. En agosto de 2022, tras una consulta con más de 50 expertos del mundo, Meta publicó su Protocolo sobre política de crisis. Esto ayudará a proporcionar una base más coherente y transparente sobre cómo Meta responde a situaciones de crisis. El 25 de enero de 2023, Meta señaló que usó el Protocolo sobre política de crisis para evaluar el entorno actual. Ejemplo de ello fue el análisis del desarrollo de las elecciones legislativas de 2022 en EE. UU., antes de su decisión sobre las cuentas del expresidente Trump.
- ◉ Muchas de nuestras propuestas han instado a Meta a proporcionar a los usuarios **mucha más transparencia** en torno a sus reglas y excepciones. En respuesta, en agosto de 2022, Meta reveló por primera vez la cantidad de **“concesiones de interés periodístico”** que aplicó al contenido infractor que consideró de interés público. Desde el 1.º de junio de 2021 hasta el 1.º de junio de 2022, Meta documentó 68 concesiones, de las cuales 13 se emitieron para publicaciones de políticos. Este tipo de transparencia ayuda a crear legitimidad y confianza con los usuarios y la sociedad civil.
- ◉ En respuesta a nuestras recomendaciones sobre sus reglas referentes a **personas y organizaciones peligrosas**, Meta inició una revisión profunda de esta área de la política. Esta revisión se centra en adoptar un enfoque basado en el riesgo para designar a personas u organizaciones como peligrosas, en el que las entidades evaluadas como de mayor riesgo se priorizarían para la aplicación. En varias de nuestras decisiones, también concluimos que la definición de “elogio” de Meta en esta política limitaba demasiado la expresión de los usuarios. En respuesta, Meta está revisando la forma de evaluar si el contenido equivale a un elogio, a un apoyo sustantivo o a una representación de una persona u organización designada.

||| TRATO JUSTO HACIA LOS USUARIOS

- ◉ En 2022, Meta **actualizó sus notificaciones de usuario sobre el contenido eliminado, luego de una solicitud del Gobierno** en virtud de la ley local. Estas indican a los usuarios que el contenido se ha restringido y explican cómo procesa Meta dichas solicitudes.
- ◉ Meta finalizó la implementación mundial de la nueva mensajería para los usuarios que indica a las personas si la **revisión manual o automatizada** ocasionó que su contenido se eliminara.
- ◉ En respuesta a nuestra decisión sobre la “afirmación sobre la existencia de una cura para el COVID-19”, **Meta creó una nueva sección en las normas comunitarias sobre la información errónea**, la cual consolida y aclara las reglas en un solo lugar.

**PUNTO
DESTACADO DE LA
RECOMENDACIÓN**

Informe independiente de derechos humanos sobre el impacto de Meta en Israel y Palestina en mayo de 2021

Tras leer en los comentarios públicos enviados para nuestro caso sobre la “publicación compartida de Al Jazeera” de 2021 afirmaciones de que Meta había eliminado desproporcionadamente las publicaciones de palestinos, instamos a Meta a contratar a una entidad independiente para que examinara si fue parcial su moderación de contenido relacionada con el conflicto en Israel y Palestina en mayo de 2021.

Meta aceptó hacerlo y encargó a la organización sin fines de lucro Business for Social Responsibility (BSR) llevar a cabo esta revisión, que se publicó en septiembre de 2022. El informe concluyó que la moderación de contenido de Meta durante el conflicto en Israel y Palestina de mayo de 2021 parecía haber repercutido de forma negativa en los derechos humanos, ya que impactó en los derechos de los usuarios palestinos a la libertad de expresión y en su capacidad de compartir información y percepciones sobre sus experiencias

como ocurrieron. Gran parte del sesgo identificado en el informe de BSR tuvo que ver con una falta de competencia y orientación en el idioma en cuestión entre los moderadores de contenido en Meta. Como Consejo, en tres decisiones distintas habíamos instado a Meta a traducir su orientación interna para los moderadores al idioma del contenido que iban a revisar. Sin embargo, Meta constantemente rechazó implementar esta recomendación con el argumento

de que la orientación en inglés es suficiente, ya que sus moderadores dominan el inglés. El informe de BSR mostró que la falta de competencias lingüísticas y conocimiento del contexto cultural entre los moderadores de Meta produjo una aplicación excesiva al contenido en árabe palestino y una aplicación insuficiente al contenido antisemita. Como hemos señalado anteriormente, la orientación solo en inglés podría provocar que los revisores pasen por alto contexto y matices en los idiomas y dialectos. Además, dado que los datos de la revisión manual se usan para capacitar a los clasificadores, este sesgo se intensifica en Facebook e Instagram.



Parece que las acciones de Meta en mayo de 2021 repercutieron de forma negativa en los derechos humanos, ya que impactaron en los derechos de los usuarios palestinos a la libertad de expresión”.

Human rights due diligence of Meta’s impacts in Israel and Palestine: informe de Business for Social Responsibility (BSR)

RECOMENDACIONES QUE REQUIEREN MÁS ATENCIÓN DE META

La implementación por parte de Meta de muchas de nuestras recomendaciones anteriores ya está mejorando las experiencias de los usuarios en Facebook e Instagram. Sin embargo, hay tres áreas en las que la empresa, hasta el momento, no ha implementado recomendaciones clave.

1. Alinear las reglas de Facebook e Instagram.

2. Traducir la orientación interna para los moderadores a los idiomas en los que moderan el contenido.

3. Proporcionar más información sobre cómo se presentan a un nivel superior las publicaciones de interés periodístico en Meta.

Alinear las reglas de Facebook e Instagram. Recomendamos por primera vez que Meta alineara las reglas de Facebook e Instagram en nuestra decisión de “síntomas de cáncer de mama y desnudos”, en enero de 2021. Aunque la empresa inicialmente se comprometió a implementar esta recomendación, en repetidas ocasiones ha retrasado la fecha límite de la implementación. Meta tampoco ha informado todavía a los usuarios, conforme a nuestra recomendación, que aplica las normas comunitarias de Facebook en Instagram y que, si el contenido se considera infractor en Facebook, también se considera infractor en Instagram.

Traducir la orientación interna para los moderadores a los idiomas en los que moderan el contenido.

En tres decisiones relacionadas con el contenido en punyabí, birmano y árabe, recomendamos que Meta tradujera sus pautas internas para los moderadores (las normas de implementación internas) a los idiomas del contenido en cuestión. A pesar de esto y de preocupaciones similares planteadas por BSR en su informe reciente, Meta ha dicho en repetidas ocasiones que “no tomará más medidas” relacionadas con esta recomendación, ya que sus moderadores hablan inglés con fluidez.

Proporcionar más información sobre cómo se presentan a un nivel superior las publicaciones de interés periodístico en Meta.

Este año, Meta reveló por primera vez la cantidad de concesiones de interés periodístico que aplicó. Sin embargo, se sabe poco acerca del proceso que usa para decidir si el contenido es de interés periodístico. En nuestra decisión sobre las “manifestaciones en Colombia” de 2021, pedimos a Meta que desarrollara y difundiera criterios claros para los revisores de contenido sobre cómo presentar a un nivel superior el contenido que infringe las reglas de Meta, pero que podría ser elegible para la concesión de interés periodístico. En su respuesta a esta recomendación, parece que Meta malinterpretó la finalidad al centrarse en el hecho de que algo *puede* presentarse a un nivel superior para recibir la concesión de interés periodístico, en lugar de centrarse en lo que solicitaba la recomendación: una descripción de *cuándo y por qué* algo se podría presentar a un nivel superior. Por lo tanto, animamos a la empresa a que priorice compartir más información en esta área.

INFORME SOBRE LA PUNTUALIDAD DE LA IMPLEMENTACIÓN DE NUESTRAS RECOMENDACIONES Y RESPUESTA A ESTAS POR PARTE DE META

- ◉ De conformidad con nuestros estatutos, Meta tiene que responder a nuestras recomendaciones públicamente en un plazo de 60 días a partir de que el Consejo publica la decisión o la opinión consultiva sobre políticas.
- ◉ En el caso de las 12 decisiones y las 2 opiniones consultivas sobre políticas publicadas en 2022, Meta respondió a nuestras recomendaciones en este plazo en todos los casos, excepto por nuestra opinión consultiva sobre políticas referida a la verificación. En este caso, debido a la gran cantidad de recomendaciones que proporcionamos (32 en total), Meta respondió a nuestras propuestas en un plazo de 90 días, en lugar de los 60 habituales.



Cronología del impacto: informar a los usuarios qué hicieron mal

En respuesta a nuestras propuestas, Meta ahora indica a más usuarios en todo el mundo *qué política específica* infringió su contenido al momento de eliminarlo.

En tres de sus primeras cinco decisiones, el Consejo recomienda que “siempre se debe notificar a los usuarios los motivos de aplicación de las normas comunitarias en su contra, incluida la regla específica que se va a aplicar”.

ENERO DE 2021

Meta desarrolla otro clasificador que puede predecir qué tipo de lenguaje que incita al odio incluye el contenido. Las personas que usan Facebook en inglés ahora reciben una mensajería más específica si el contenido se elimina debido a lenguaje que incita al odio.

JULIO DE 2021

Meta responde y dice que está evaluando la viabilidad de esta recomendación.

FEBRERO DE 2021

NOVIEMBRE DE 2021

Meta comienza a probar una mensajería más específica para las infracciones de uso de lenguaje que incita al odio en Facebook en idiomas que no son el inglés y empieza por el árabe, el español y el portugués. También inicia pruebas de esta mensajería en Instagram.

Meta comienza a probar una mensajería más específica para los usuarios sobre contenido eliminado por infringir su política de bullying y acoso.

MARZO DE 2022

**MARZO DE
2022**

Meta afirma que las notificaciones más específicas generaron “un aumento estadísticamente significativo de las percepciones de su transparencia y legitimidad en varios mercados”.

**MAYO DE
2022**

Meta afirma que su investigación ha demostrado que, para garantizar que las personas se sientan escuchadas, su proceso de revisión de contenido debe funcionar más como un diálogo que promueve un entendimiento mutuo.

**AGOSTO DE
2022**

Meta lanza una mensajería nueva que explica a las personas exactamente qué política ocasionó que se tomaran medidas de aplicación. Esta información está disponible en inglés en todo el mundo (próximamente habrá traducciones en otros idiomas).

**NOVIEMBRE
DE 2022**

Meta afirma que esta mensajería nueva se aplicará a la gran mayoría de los tipos de infracción en Facebook para finales de 2022 y se extenderá a Instagram en 2023. También evaluará la viabilidad de que las notificaciones sean aún más específicas.

Implementación de nuestras recomendaciones por parte de Meta

Para garantizar que Meta cumpla sus compromisos, supervisamos su avance en la implementación de nuestras recomendaciones. Para ello, analizamos si se cumplieron los criterios de una recomendación concreta. Medimos la implementación de Meta de acuerdo con siete categorías y actualizamos nuestras evaluaciones cada tres meses.



“Implementación demostrada mediante información publicada”

Meta ha proporcionado datos suficientes para que el Consejo verifique que se haya implementado la recomendación.



“Implementación parcial demostrada mediante información publicada”

Meta ha implementado un componente central de la recomendación y ha proporcionado datos suficientes al Consejo para verificar esto.



“Progreso informado”

Meta se ha comprometido a implementar esta recomendación, pero aún no ha completado todas las acciones necesarias.



“Meta informó la implementación o describió la recomendación como un trabajo que Meta ya hace, pero no publicó la información para demostrar la implementación”

Meta dice que implementó esta recomendación, pero no ha proporcionado pruebas suficientes para que podamos verificarlo.



“Recomendación rechazada después de la evaluación de viabilidad”

Meta se comprometió a implementar la recomendación y, luego, decidió rechazar su implementación después de proporcionar información sobre la decisión.



“Recomendación omitida, rechazada o reformulada”

Meta no tomará más medidas relacionadas con nuestra propuesta.










“En espera de la primera respuesta de Meta”

El Consejo hizo la recomendación, pero Meta aún no ha respondido públicamente, ya que tiene 60 días a partir de la publicación para responder.

Este enfoque basado en datos significa que nuestra evaluación de si Meta ha implementado una recomendación puede diferir de los informes de la empresa. Sin embargo, creemos que este tipo de validación independiente es crucial para asegurar la responsabilidad y para garantizar que los usuarios sientan el impacto de nuestras recomendaciones.

La tabla en esta página brinda un desglose de nuestra evaluación de la implementación por parte de Meta de las 191 recomendaciones que hicimos hasta principios de abril de 2023, cuando se finalizó este informe. También publicamos un documento adjunto a este informe anual que indica nuestras 191 recomendaciones en su totalidad, lo que proporciona nuestra evaluación de la respuesta y la implementación de Meta. Esto refleja los informes trimestrales de Meta sobre el Consejo hasta el cuarto trimestre de 2022.

CATEGORÍA DE IMPLEMENTACIÓN	N.º DE RECOMENDACIONES
 Implementación demostrada mediante información publicada	27
 Implementación parcial demostrada mediante información publicada	14
 Progreso informado	84
 Meta informó la implementación o describió la recomendación como un trabajo que Meta ya hace, pero no publicó la información para demostrar la implementación	29
 Recomendación rechazada después de la evaluación de viabilidad	10
 Recomendación omitida, rechazada o reformulada	23
 En espera de la primera respuesta de Meta	4
NÚMERO TOTAL DE RECOMENDACIONES	
	191

Selección de casos

Se presentaron

1.290.942 casos al Consejo en 2022

Lo que supone un aumento de aproximadamente un cuarto en comparación con 2021



EN PROMEDIO,
el Consejo recibió un caso cada 24 segundos en 2022

Más de dos tercios de las apelaciones de usuarios para restaurar contenido se referían a solo dos normas comunitarias:



violencia e incitación



lenguaje que incita al odio

EN CASI DOS TERCIOS DE LOS CASOS QUE SELECCIONÓ EL CONSEJO EN 2022,

Meta identificó como incorrecta su decisión original sobre el contenido

Información general

Uno de nuestros retos más grandes es decidir qué casos revisar, ya que nos encontramos con cientos de millones de publicaciones compartidas en Facebook e Instagram todos los días y contamos con decenas de miles de moderadores de contenido que toman decisiones en una fracción de segundo sobre qué contenido se queda o se elimina. En 2022, recibimos aproximadamente 1,3 millones de solicitudes de usuarios para revisar de manera independiente las decisiones de moderación de contenido de Meta. Meta también remitió 21 casos al Consejo. En 2022, recibimos, en promedio, 3.537 casos al día.

Nuestro objetivo es centrarnos en casos que suponen cuestiones difíciles y relevantes, y que reflejan una gran variedad de regiones geográficas y ámbitos. Mediante la selección de casos que nos permiten analizar las políticas que más importan a los usuarios y la aplicación de una revisión atenta y fundamentada, podemos mejorar las políticas para todos los usuarios, no solo para aquellos cuyos casos se atienden. Por ejemplo, seleccionar el caso de la “caricatura sobre la policía colombiana” nos dio la oportunidad de hacer recomendaciones sobre los bancos de coincidencias de contenido multimedia de Meta, que pueden eliminar automáticamente las imágenes que podrían infringir las reglas de Meta. De manera similar, el caso sobre “música drill británica” brindó una oportunidad para hacer propuestas sobre cómo debe responder Meta a las solicitudes de organismos policiales en todo el mundo.

Para abordar las cuestiones que son relevantes para las personas en todo el mundo, continuamos seleccionando casos de distintas regiones en todo el mundo. También elegimos los casos que son de importancia fundamental para el discurso público, como el caso del “poema ruso”, en el que se examinaron las cuestiones de expresión en el contexto de la invasión rusa a Ucrania, y el caso de “reapropiación de términos árabes” sobre el uso de términos despectivos que afectan a las personas de la comunidad LGBTQIA+.

“Intentamos elegir los casos que abordan los problemas que suele enfrentar la empresa o que tienen un impacto real en estos”.

Michael McConnell
COPRESIDENTE DEL CONSEJO
ASESOR DE CONTENIDO



LECCIONES APRENDIDAS

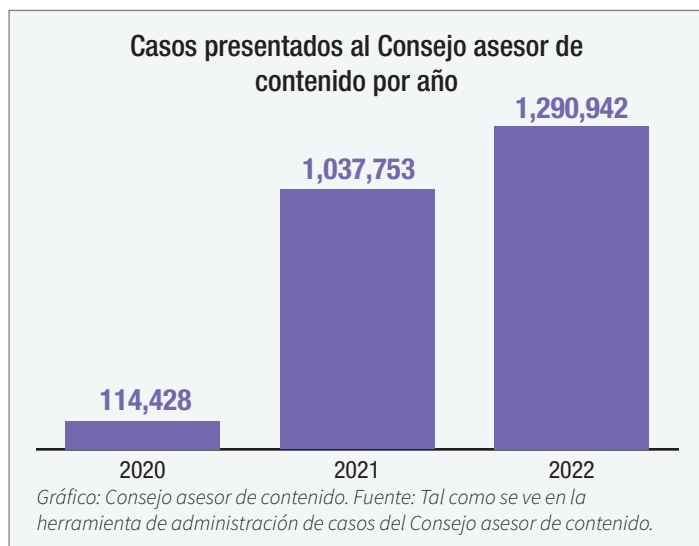
A finales de 2022, al preparar nuestra decisión sobre “música drill británica” y la opinión consultiva sobre políticas referida al programa de verificación de Meta, identificamos un problema con el proceso de apelación de la empresa: para determinadas apelaciones presentadas a un nivel superior, no se había ofrecido a los usuarios la oportunidad de apelar ante el Consejo, a pesar de que dichos casos eran elegibles para su revisión de conformidad con nuestros estatutos. Esto planteó importantes preocupaciones sobre la transparencia y el derecho de los usuarios a la enmienda. En nuestra decisión sobre “música drill británica” y nuestra opinión consultiva sobre políticas referida al programa de verificación de Meta, instamos a la empresa a rectificar este error.

En enero de 2023, Meta respondió y dijo que las personas en la UE, el Reino Unido y la India pronto podrían apelar ante Meta y ante el Consejo las decisiones sobre contenidos elegibles que se hayan tomado al presentar el caso a un nivel superior. Los equipos internos de especialistas de Meta se encargan de las decisiones tomadas “al presentar el caso a un nivel superior”, en lugar de los contratistas externos. En el caso de los usuarios de Facebook e Instagram en otros países, Meta anunció planes para desarrollar una vía alternativa que permita a los usuarios apelar directamente ante nosotros las decisiones de eliminación elegibles que se hayan tomado al presentar el caso a un nivel superior y que no puedan apelarse internamente. La empresa indicó que esperaba implementar esta solución para la segunda mitad de 2023.

Casos presentados al Consejo

En 2022, la cantidad de apelaciones que recibimos de los usuarios aumentó alrededor de un cuarto, de aproximadamente 1 millón en 2021 a aproximadamente 1,3 millones en 2022. Aunque en el primer y segundo trimestre de 2022 se vieron cifras de apelaciones más altas que en los trimestres correspondientes en 2021, en el tercer y cuarto trimestre de 2022 se vieron menos apelaciones que un año atrás.

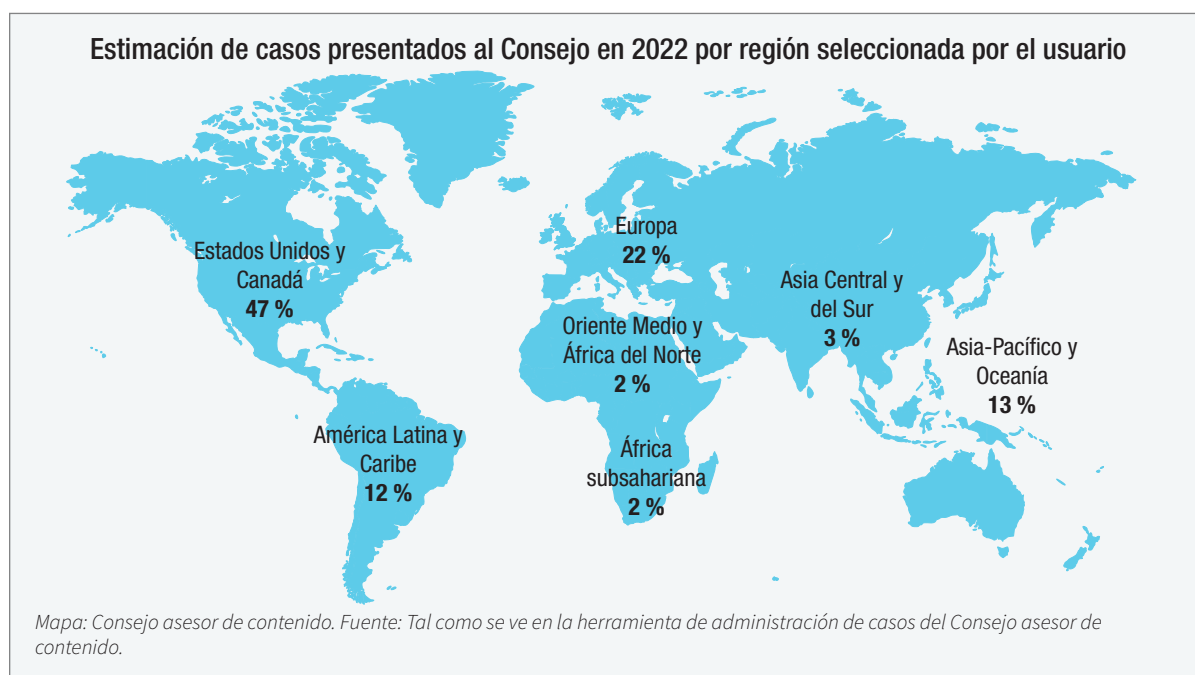
En total, entre octubre de 2020, cuando empezamos a aceptar casos, y diciembre de 2022, recibimos más de 2,4 millones de apelaciones, lo que refleja la demanda continua de los usuarios para apelar ante un órgano independiente las decisiones de moderación de contenido de Meta. Aunque solo podemos revisar una cantidad pequeña



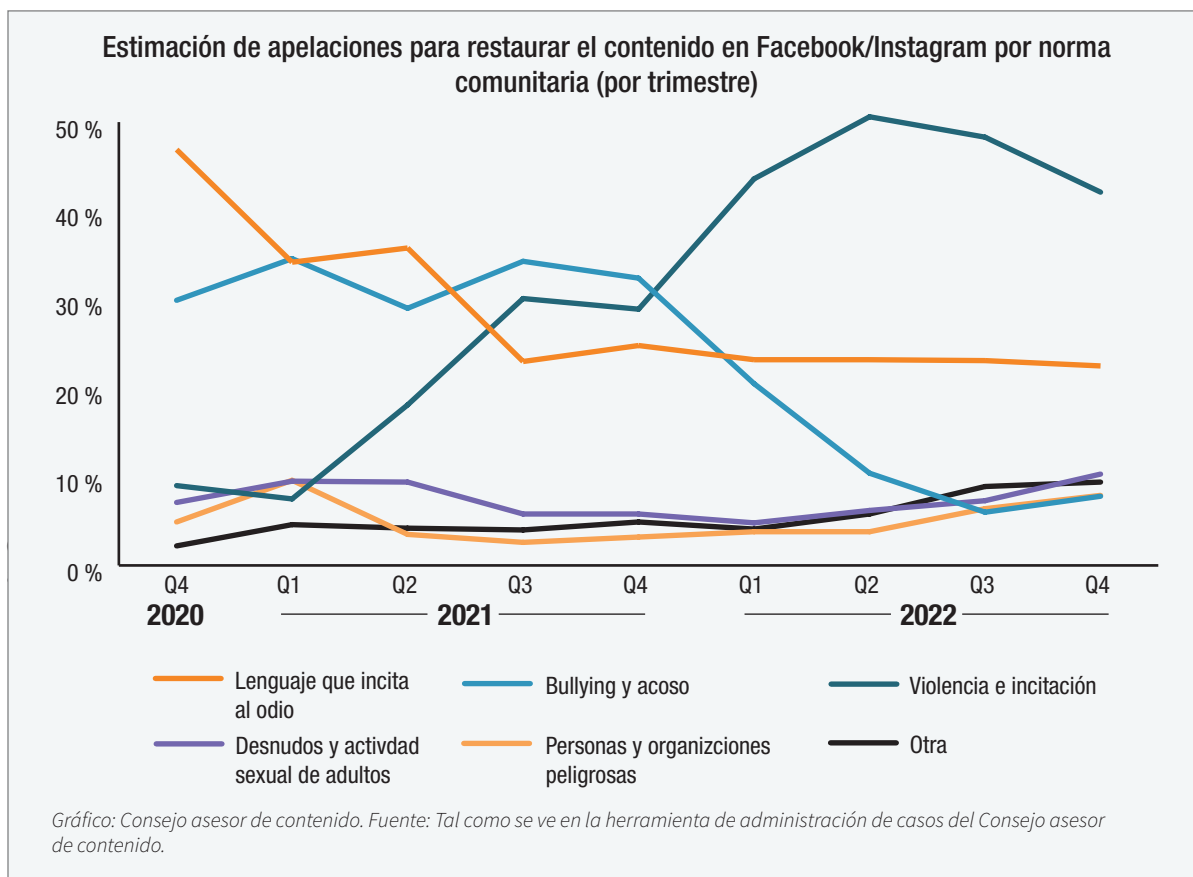
de casos, seguimos seleccionando los que plantean problemas subyacentes a los que se enfrentan grandes cantidades de usuarios alrededor del mundo y hacemos recomendaciones para abordarlos.

En 2022 se vio un aumento significativo del porcentaje de casos sobre contenido en Instagram. Aunque estos constituyeron solo el 1 % de las apelaciones en cada trimestre hasta el primer trimestre de 2022, para el cuarto trimestre de 2022, este porcentaje había aumentado a más del 10 % de apelaciones.

En 2022, el desglose de casos por región seleccionada por el usuario permaneció prácticamente igual desde octubre de 2020 hasta diciembre de 2021, con más de dos tercios de todas las apelaciones procedentes de EE. UU., Canadá y Europa. El porcentaje de casos procedentes de EE. UU. y Canadá descendió ligeramente del 49 % en 2020/21 al 47 % en 2022. El porcentaje de casos de Europa aumentó del 20 % en 2020/21 al 22 % en 2022. Hubo un aumento notorio en el porcentaje de casos de Asia-Pacífico y Oceanía, que aumentó del 9 % en 2020/21 al 13 % en 2022. El 12 % de los casos provino de Latinoamérica y el Caribe en 2022, seguido del 3 % de Asia Central y Asia del Sur, el 2 % de Oriente Medio y África del Norte, y el 2 % de África subsahariana. Reconocemos que estas cifras no reflejan la difusión de los usuarios de Facebook e Instagram en todo el mundo, ni la distribución real de los problemas de moderación de contenido en todo el mundo. En 2023, aumentaremos la divulgación y la participación en regiones fuera de EE. UU. y Europa para concientizar sobre nuestro trabajo y animar a los usuarios de estas regiones a presentar casos al Consejo.



En 2022, más de dos tercios (69 %) de las apelaciones de usuarios para restaurar contenido se relacionaron con solo dos normas comunitarias: violencia e incitación (el 46 % de las apelaciones para restaurar contenido) y lenguaje que incita al odio (el 23 % de las apelaciones para restaurar contenido). Estas cifras solo se refieren a las apelaciones de usuarios para *restaurar* contenido que Meta consideró que infringió sus reglas (lo cual representó el 92 % de las apelaciones en 2022) y no a las apelaciones para *eliminar* el contenido de otras personas que aún se encuentra publicado en Facebook o Instagram (lo cual representó el 8 % de las apelaciones en 2022), ya que supuestamente este último no infringió una norma comunitaria.



Tal como se muestra en el gráfico, en 2022 aumentó el porcentaje de las apelaciones de usuarios para restaurar el contenido eliminado de conformidad con la norma comunitaria de violencia e incitación de Meta; este alcanzó un máximo del 51 % en el segundo trimestre de 2022. Sin embargo, las apelaciones de usuarios para restaurar el contenido eliminado de conformidad con su norma de bullying y acoso descendieron de aproximadamente el 30 % en el cuarto trimestre de 2021 a menos del 10 % en la segunda mitad de 2022.

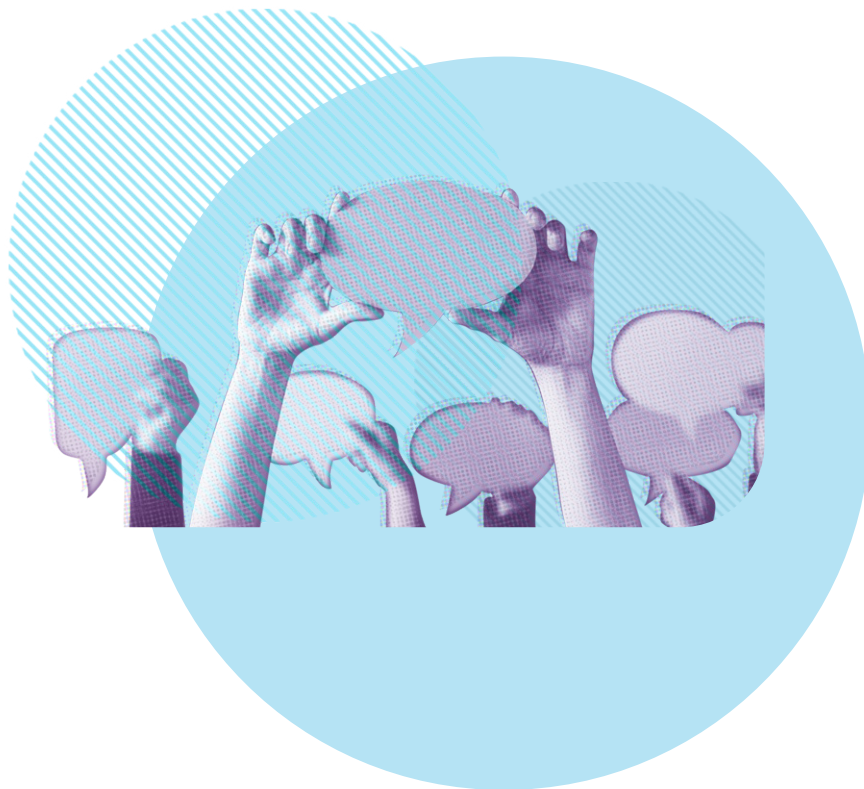
Los informes de aplicación de las normas comunitarias de Meta indicaron un aumento similar de la cantidad de contenido de violencia e incitación procesado en Facebook en 2022: aumentó de 12,7 millones en el cuarto trimestre de 2021 a 21,7 millones en el primer trimestre de 2022 y 19,3 millones en el segundo trimestre de 2022. Esto puede explicar en parte el aumento de las apelaciones ante el Consejo. El informe de Meta también mostró que, en el tercer trimestre de 2022, la cantidad de contenido de bullying y acoso que Meta procesó en Facebook disminuyó a 6,6 millones, su nivel más bajo en casi 2 años. Esto puede explicar en parte la reducción del porcentaje de apelaciones ante el Consejo para restaurar el contenido eliminado de conformidad con esta norma comunitaria.

Casos que consideró el Comité de selección de casos

Después de que el Comité de selección de casos del Consejo selecciona los casos para que el Consejo los revise, Meta a veces determina que su decisión original sobre un contenido fue incorrecta.

Meta consideró que su decisión original en 32 de 50 casos seleccionados en 2022 (64 %) fue incorrecta, en comparación con 51 de 130 casos seleccionados (39 %) en 2020/21. Esto representa un aumento de 25 puntos porcentuales.

Aunque esto es solo una pequeña muestra, y el Consejo busca a propósito los casos más complejos y difíciles, es preocupante que en casi dos tercios de los casos seleccionados Meta concluyera que su decisión original fue incorrecta. Esto plantea cuestiones más generales acerca de la precisión de la moderación de contenido de Meta y el proceso de apelación que aplica Meta antes de que los casos lleguen al Consejo. En 2023, comenzaremos a publicar las “decisiones sumarias”. Estas examinarán los casos que no seleccionamos para su revisión completa, pero que ocasionan que Meta revierta su medida de aplicación original.



Decisiones de casos y opiniones consultivas sobre políticas

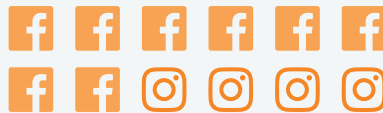
Se publicaron

12 decisiones en 2022

9 de ellas anularon la decisión de Meta



7 casos de los usuarios
5 casos de Meta



8 casos de Facebook
4 casos de Instagram

Se publicaron las primeras opiniones consultivas sobre políticas del Consejo con respecto a y



compartir información privada



el programa de verificación de Meta



Se pudo indicar a Meta que deje o restaure el contenido con una pantalla de advertencia

Se hicieron 356 preguntas a Meta como parte de nuestra revisión de casos en 2022



De las cuales Meta respondió completamente el 86 %

Información general

Las decisiones que tomamos en respuesta a las apelaciones de usuarios y las remisiones de Meta son la base del trabajo del Consejo. En 2022, también publicamos nuestras primeras opiniones consultivas sobre políticas referentes a compartir información de residencia privada y el programa de verificación de Meta.

DECISIONES DE CASOS

Nuestras decisiones de casos examinan si las elecciones de Meta con respecto a eliminar o dejar el contenido están alineadas con las reglas, los valores y los compromisos asumidos en materia de derechos humanos de la empresa. Nuestras decisiones son vinculantes para Meta, y la empresa debe implementarlas en un plazo de siete días a partir de la publicación, a menos que hacerlo implique infringir la ley. En 2022, publicamos 12 decisiones de casos, las cuales anulaban la decisión de Meta sobre el contenido en cuestión el 75 % de las veces, ligeramente por encima del 70 % en 2021.

Hicimos también varios cambios en nuestro proceso de revisión para aumentar nuestro impacto. En julio de 2022, anunciamos que consideraríamos dos casos sobre identidad de género y desnudos en el texto de una decisión por primera vez. Considerar los casos en conjunto nos ayudará a explorar una mayor variedad de experiencias de los usuarios y a comparar las diferencias entre publicaciones al límite. En octubre de 2022, también ampliamos nuestro ámbito y pudimos tomar decisiones vinculantes para aplicar una pantalla de advertencia que marca el contenido como “perturbador” o “sensible” al dejar o restaurar contenido que cumple los requisitos. En noviembre, publicamos nuestra decisión sobre “música drill británica”. Como parte de esta decisión, usamos una solicitud de libertad de la información para obtener nuevos detalles sobre cómo la Policía Metropolitana de Londres hace solicitudes a las empresas de redes sociales para eliminar contenido.

OPINIONES CONSULTIVAS SOBRE POLÍTICAS

Aparte de revisar los casos individuales para eliminar o restaurar el contenido, también podemos aceptar opiniones consultivas sobre políticas de Meta. Mediante estas, proporcionamos a la empresa recomendaciones detalladas sobre los cambios que debe hacer Meta en sus políticas sobre un tema concreto. Creemos que estas recomendaciones también pueden ser útiles para otras empresas que, a menudo, enfrentan problemas de moderación de contenido similares a los que enfrenta Meta.

En 2022, emitimos nuestras primeras opiniones consultivas sobre políticas respecto a compartir información de residencia privada y el programa de verificación de Meta. Nuestra opinión consultiva sobre políticas en cuanto a la verificación compartió un nivel de detalle sin precedentes sobre el programa, e hicimos 32 recomendaciones a la empresa.

CÓMO RESPONDIÓ META A NUESTRAS PREGUNTAS

Para ayudar con nuestras decisiones y opiniones consultivas sobre políticas, enviamos preguntas a la empresa antes de deliberar. Para las 12 decisiones y las 2 opiniones consultivas sobre políticas que publicamos en 2022, hicimos 356 preguntas a Meta. Meta respondió 306 de estas preguntas completamente, 37 parcialmente, y no respondió 13.

El porcentaje de preguntas que Meta respondió completamente en 2022 (86 %) fue idéntico al de 2021. Sin embargo, el porcentaje de preguntas que Meta no respondió disminuyó del 6 % (19 de 313 preguntas) en 2021 al 4 % (13 de 356 preguntas) en 2022. Creemos que las respuestas de Meta en 2022 fueron más integrales que en 2021. Cuando Meta no pudo responder una pregunta, explicó el motivo más a menudo que en 2021.

LECCIONES APRENDIDAS

En 2022, nuestros estatutos establecían como plazo de publicación de las decisiones los 90 días siguientes al anuncio del caso en nuestro sitio web. Los estatutos también estipulaban excepciones a este plazo, por ejemplo, en circunstancias excepcionales o en caso de incidentes técnicos u operativos. En 2022, hubo más prórrogas excepcionales que casos resueltos a tiempo. En algunos casos, se prolongaron más de lo previsto las negociaciones con Meta sobre la cantidad de información que podíamos incluir en nuestra decisión final, proporcionada en un principio por la empresa de manera confidencial. En otros casos, los problemas de traducción retrasaron la publicación. En otras ocasiones, las limitaciones de programación del Consejo y otros problemas nos impidieron hacer la publicación en un plazo de 90 días. Nuestro objetivo sigue siendo resolver los casos en un plazo de 90 días (tal como se refleja en la actualización de nuestros estatutos de febrero de 2023) y ser transparentes en cuanto a la cronología de los casos.

Aunque el Consejo publicó dos opiniones consultivas sobre políticas en 2022, en mayo de ese año también anunciamos que Meta había informado al Consejo que la empresa retiraría una solicitud anterior de orientación sobre políticas en cuanto a cuestiones de moderación de contenido relacionadas con la actual guerra de Rusia con Ucrania. Al tomar esta medida, la empresa citó preocupaciones concretas y continuas de seguridad y protección. Aunque comprendemos estas preocupaciones, consideramos que la solicitud planteaba cuestiones importantes y nos decepcionó la decisión de Meta de retirarla. Sin embargo, más tarde seleccionamos y publicamos nuestra decisión sobre el “poema ruso”, que abarcaba varias cuestiones de moderación de contenido relacionadas con la invasión rusa a Ucrania.

Decisiones y opiniones consultivas sobre políticas emitidas en 2022

DECISIONES DE CASOS

	PLATAFORMA	FUENTE	NORMA COMUNITARIA	PAÍSES	DECISIÓN DEL CONSEJO
Dudas sobre cómo pedir Adderall® N.º de caso: 2021-015-FB-UA		Usuario	Bienes y servicios restringidos	Estados Unidos	Se anuló la decisión de Meta de eliminar el contenido
Denuncia por parte de un periodista sueco de violencia sexual contra menores N.º de caso: 2021-016-FB-FBR		Meta	Desnudos y actividad sexual de adultos	Suecia	Se anuló la decisión de Meta de eliminar el contenido
Dibujos animados de Knin N.º de caso: 2022-001-FB-UA		Usuario	Lenguaje que incita al odio	Croacia	Se anuló la decisión de Meta de dejar el contenido
Video gráfico de Sudán N.º de caso: 2022-002-FB-MR		Meta	Contenido violento y gráfico	Sudán	Se confirmó la decisión de Meta de dejar el contenido
Reapropiación de términos árabes N.º de caso: 2022-003-IG-UA		Usuario	Lenguaje que incita al odio	Marruecos, Egipto, Líbano	Se anuló la decisión de Meta de eliminar el contenido
Caricatura sobre la policía colombiana N.º de caso: 2022-004-FB-UA		Usuario	Personas y organizaciones peligrosas	Colombia	Se anuló la decisión de Meta de eliminar el contenido
Mención de los talibanes en los informes de noticias N.º de caso: 2022-005-FB-UA		Usuario	Personas y organizaciones peligrosas	Afganistán	Se anuló la decisión de Meta de eliminar el contenido
Consejo de comunicaciones de Tigray N.º de caso: 2022-006-FB-MR		Meta	Violencia e incitación	Etiopía	Se confirmó la decisión de Meta de eliminar el contenido
Música drill británica N.º de caso: 2022-007-IG-MR		Meta	Violencia e incitación	Reino Unido	Se anuló la decisión de Meta de eliminar el contenido
Poema ruso N.º de caso: 2022-008-FB-UA		Usuario	Lenguaje que incita al odio	Letonia, Ucrania, Rusia	Se anuló la decisión de Meta de eliminar el contenido
Video tras el ataque en una iglesia de Nigeria N.º de caso: 2022-011-IG-UA		Usuario	Contenido violento y gráfico	Nigeria	Se anuló la decisión de Meta de eliminar el contenido
Video sobre acoso sexual en la India N.º de caso: 2022-012-IG-MR		Meta	Explotación sexual de adultos	Nigeria	Se confirmó la decisión de Meta de dejar el contenido

OPINIONES CONSULTIVAS SOBRE POLÍTICAS

Uso compartido de información de residencia privada
N.º de caso: PAO-2021-01

Programa de verificación de Meta
N.º de caso: PAO-2021-02

Resúmenes de decisiones y opiniones consultivas sobre políticas

En octubre de 2022, anunciamos siete prioridades estratégicas en función de un análisis profundo de las cuestiones y preocupaciones que más se plantearon en las apelaciones de los usuarios. En esta sección, se agrupan las decisiones y las opiniones consultivas sobre políticas que publicamos en 2022 de acuerdo con estas prioridades.

Elecciones y espacio cívico



Mención de los talibanes en los informes de noticias

ANULADA

En enero de 2022, un popular periódico en urdu de la India publicó en su página de Facebook un anuncio de un portavoz oficial del régimen talibán en Afganistán, en el que se afirmaba que las escuelas y las universidades para mujeres y niñas volverían a abrir en dos meses. Meta eliminó la publicación por infringir la norma de Facebook sobre personas y organizaciones peligrosas, la cual prohíbe el “elogio” de entidades, incluidas las organizaciones terroristas, que se considera que “ocasionan daños graves en la vida real”. Meta también impuso “faltas” al administrador de la página y limitó el acceso de la cuenta a algunas funciones de Facebook.

Después de que el usuario apelara la decisión de Meta, una segunda revisión manual confirmó la evaluación inicial de la empresa que determinaba que la publicación era infractora. Sin embargo, después de seleccionar el caso para su revisión, Meta reconoció que la decisión original de eliminar el contenido había sido un “error del sistema de cumplimiento de normas”, restauró el contenido, eliminó las “faltas” y canceló las restricciones de la cuenta. A pesar del cambio de rumbo de Meta, anulamos la decisión original de Meta de eliminar este anuncio positivo del régimen talibán. Citamos una excepción específica de la norma comunitaria sobre personas y organizaciones peligrosas que permite la publicación de informes de interés periodístico sobre las actividades de grupos terroristas. También observamos que no estaba clara la relación entre la concesión de “informes” de la política sobre personas y organizaciones peligrosas y la concesión de interés periodístico general.

“ El Consejo decidió que, aunque los talibanes son una organización peligrosa, los informes de noticias acerca de estos no se deben restringir, dado el interés público en cuestión”.

Paolo Carozza

MIEMBRO DEL CONSEJO ASESOR DE CONTENIDO



¿Qué normas comunitarias examinaron más nuestras decisiones en 2022?

Lenguaje que incita al odio

🗨️ 🗨️ 🗨️ 3 decisiones

Personas y organizaciones peligrosas

👤 👤 2 decisiones

Violencia e incitación

📣 📣 2 decisiones

Contenido violento y gráfico

🚫 🚫 2 decisiones



Denuncia por parte de un periodista sueco de violencia sexual contra menores

ANULADA

En agosto de 2019, un usuario de Suecia publicó en su página de Facebook una foto de archivo de una niña pequeña sentada con las manos en la cabeza, ocultando su rostro. La foto estaba acompañada de descripciones detalladas de las violaciones de dos menores anónimos. La publicación daba las edades de dos perpetradores condenados anónimos e incluía detalles gráficos del impacto nocivo que tuvo el delito en una víctima. También contenía citas atribuidas a un perpetrador en las que se jactaba de la violación y se refería a una de las víctimas con términos sexualmente explícitos. En septiembre de 2021, después de dejarla en la plataforma durante dos años, Meta eliminó la publicación por infringir la política sobre explotación sexual, maltrato y desnudos de menores.

Nuestra decisión sostuvo que era incorrecta la decisión de Meta de eliminar la publicación. Comunicar una descripción clínica de las secuelas de una violación y la declaración sexualmente explícita del perpetrador sobre esta no equivalía a un mensaje que “explotara sexualmente a niños o representara a un menor en un contexto de carácter sexual”. El contexto general de la publicación confirmaba con claridad que la intención del usuario había sido informar un asunto de evidente interés público, además de condenar la explotación sexual de un menor. Entre nuestras recomendaciones en este caso, instamos a Meta a distinguir de manera clara entre el contenido que apoya o promueve la explotación sexual de menores y el contenido que concientiza sobre ella.

Situaciones de crisis y conflicto



Consejo de comunicaciones de Tigray

CONFIRMADA

A principios de 2022, Meta remitió un caso sobre una publicación que apareció en la página oficial de Facebook del Consejo de comunicaciones del estado regional de Tigray, organismo gestionado por las autoridades de una provincia etíope. En la publicación, hecha durante el actual conflicto entre las fuerzas federales y tigrayanas en la región, se hacía referencia a las recientes pérdidas sufridas por las fuerzas federales y se instaba al ejército nacional a “abrir fuego” contra el “grupo de Abiy Ahmed”, en referencia al primer ministro de Etiopía. Continuaba afirmando que, si las fuerzas federales se negaban a acatar esta advertencia, morirían.

Después de que varios usuarios reportaran la publicación y los sistemas automatizados de Meta la identificaran, dos revisores que hablaban amárico decidieron que esta no infringía las políticas de Meta. Dos días después, tras presentar la publicación para que la revisaran

expertos en un Centro de operaciones de productos relativos a la integridad (IPOC) creado por la empresa para moderar los contenidos surgidos del conflicto, Meta revirtió su decisión original y eliminó la publicación por infringir la política de Facebook sobre violencia e incitación.

Aunque reconocemos que Meta ha tomado medidas positivas, como el establecimiento del sistema IPOC, para supervisar el contenido y detectar abusos en situaciones de conflicto de alto riesgo, nuestra decisión resaltó que el contenido infractor había permanecido en la plataforma durante 2 días y se había visto más de 300.000 veces antes de su eliminación. Como mínimo, eso por sí solo subrayaba el hecho de que los IPOC “no están diseñados para ser una solución sostenible a largo plazo a fin de lidiar con un conflicto de años”. En respuesta, recomendamos que “es posible que Meta deba invertir en un mecanismo más duradero” para cumplir con sus responsabilidades en materia de derechos humanos en zonas de conflicto y situaciones de crisis.



Video gráfico de Sudán

CONFIRMADA

En las últimas semanas de 2021, Meta remitió un caso sobre un video gráfico publicado en una página de perfil de Facebook tras el golpe militar ocurrido en octubre de 2021 en Sudán. El video mostraba a una persona tendida junto a un coche con una herida importante en la cabeza y un ojo visiblemente desprendido. De fondo, se oían voces que decían en árabe que alguien había sido golpeado y abandonado en la calle. Un texto, también en árabe, pedía a la población que no confiara en el ejército y citaba una prolongada serie de abusos.

Después de que Meta eliminara la publicación por infringir la norma comunitaria de Facebook sobre contenido violento y gráfico, la empresa emitió una “concesión de interés periodístico”. Tras un retraso de casi cinco semanas, Meta restauró el contenido con una pantalla de advertencia que restringía su visualización.

Nuestra decisión confirmó la decisión de Meta de restaurar el contenido con una pantalla de advertencia. También sostuvimos que aplicar una “concesión de interés periodístico” no era un medio eficaz para moderar el contenido gráfico compartido en Facebook a escala. Para subrayar esta conclusión, citamos la admisión de Meta de que había emitido 17 concesiones de interés periodístico en relación con la política sobre contenido violento y gráfico durante los tres primeros trimestres de 2021. En cambio, había eliminado más de 90 millones de contenidos por infringir la misma norma durante el mismo período. Por ello, recomendamos que Meta revisara su norma comunitaria sobre contenido violento y gráfico para permitir que se compartan videos gráficos cuando “tengan la intención de generar conciencia o documentar abusos”.



Poema ruso

ANULADA

En abril de 2022, un usuario de Facebook en Letonia publicó una imagen de un cadáver tendido en una calle, que Meta confirmó que era de una persona tiroteada en Ucrania. En el texto en ruso que acompañaba a la publicación, se argumentaba que, durante la Segunda Guerra Mundial, las presuntas atrocidades que los soldados soviéticos cometieron en Alemania se habían excusado como venganza por los crímenes que los soldados nazis supuestamente habían cometido en la URSS. Mencionando presuntas atrocidades contemporáneas que cometieron los soldados rusos en Ucrania, citaba un poema del poeta soviético Konstantin Simonov: “Maten al fascista... ¡Mátenlo!”.

Después de que Meta eliminó la publicación por infringir su norma comunitaria sobre lenguaje que incita al odio, seleccionamos el caso para su revisión. Meta posteriormente restauró la publicación con una pantalla de advertencia. En nuestra decisión, establecimos una distinción entre el contenido que podría haber hecho referencia a la *nacionalidad* de los soldados rusos, que habría sido infractor, y el contenido que establece un *paralelismo histórico* con las acciones que los nazis llevaron a cabo, que no es infractor. Concluimos que la cita de versos del poema de la era soviética era una referencia artística y cultural empleada para *describir*, no para *fomentar*, una mentalidad. Si bien reconocemos las complejidades de evaluar el lenguaje que incita a la violencia en situaciones de conflicto en las que el derecho internacional permite apuntar a los combatientes, recomendamos que Meta revisara sus políticas para tener en cuenta de manera clara un contexto de “intervenciones militares ilegales”.

Lenguaje que incita al odio contra grupos marginados



Dibujos animados de Knin

ANULADA

En diciembre de 2021, una página pública de Facebook publicó un video basado en el dibujo animado de Disney “El flautista de Hamelín” que contenía un texto, en croata, que Meta tradujo como “El flautista de Čavoglave” (un pueblo de Croacia) “y las ratas de Knin” (una ciudad de Croacia). El narrador relataba que, cuando una población de ratas decidió que quería vivir en un “país exclusivo de ratas”, empezó a acosar y perseguir a las personas, lo que le permitió apoderarse de él. Sin embargo, después de que apareciera un flautista de Čavoglave y tocara una melodía con su “flauta mágica”, atrajo a las ratas y las sacó de la ciudad para meterlas en un tractor, que desapareció. El narrador concluyó: “Las ratas desaparecieron para siempre de estas tierras... [y] todos vivieron felices por siempre”.

Incluso después de que los usuarios reportaran el contenido casi 400 veces, Meta no lo eliminó. Después de apelar el caso ante el Consejo, Meta llevó a cabo otra revisión manual, que también determinó que el contenido no era infractor. Una vez que seleccionamos el caso para su revisión, Meta revirtió esa decisión y determinó que, si bien la publicación no infringió la *letra*, sí infringió el *espíritu* de la política sobre lenguaje que incita al odio, y la eliminó. Durante la redacción de una explicación para el Consejo, Meta volvió a cambiar de opinión y esta vez determinó que la publicación infringió, en efecto, la *letra* de la política.

Al concluir que el contenido sí infringió las normas comunitarias sobre lenguaje que incita al odio y violencia e incitación, citamos comentarios que contenían referencias a una operación militar croata de 1995, la “Operación Tormenta”. Esta operación provocó el desplazamiento forzoso, la ejecución y la desaparición de civiles de etnia serbia en Croacia. En el contexto de esas referencias, expresamos nuestra preocupación por el hecho de que casi 40 moderadores de habla croata decidieron que el contenido no era infractor porque pensaban que la norma exigía una comparación despectiva *explícita* entre la etnia serbia y las ratas. Recomendamos que Meta aclarara la norma comunitaria sobre lenguaje que incita al odio para expresar de manera clara que la política prohíbe referencias hostiles *implícitas* y *explícitas* a grupos protegidos.



Reapropiación de términos árabes

ANULADA

En noviembre de 2021, una cuenta pública de Instagram descrita como un espacio para “debatir narrativas queer en la cultura árabe” publicó imágenes con un texto en árabe que explicaba que cada imagen contenía una palabra utilizada en el mundo de habla árabe para denigrar a los hombres con “maneras afeminadas”. El usuario afirmó que su intención era “reapropiar [el] poder de términos tan hirientes”.

Después de que Meta eliminara el contenido por infringir su política sobre lenguaje que incita al odio, lo restaurara tras la apelación del usuario y lo volviera a eliminar luego de que otro usuario lo reportara, seleccionamos el caso. Meta presentó el contenido a un nivel superior para realizar otra revisión, en la que se concluyó que el contenido no infringía la política. Meta la restauró posteriormente y afirmó que sus decisiones iniciales se habían basado únicamente en las revisiones de las imágenes que contenían los términos despectivos, no en el contenido en sí ni en la explicación de la intención.



Meta tiene valores y normas, pero donde falla es en la aplicación”.

Endy Bayuni
MIEMBRO DEL CONSEJO
ASESOR DE CONTENIDO



Al considerar que la eliminación del contenido no se alineaba con la política sobre lenguaje que incita al odio, señalamos que el contenido entra en el ámbito de una excepción que permite compartir el lenguaje “utilizado de forma autorreferencial o con ánimo de empoderamiento”. Se permite citar lenguaje que incita al odio con la intención de condenarlo o generar conciencia. La toma de decisiones contradictorias de Meta en este caso reflejaba, en nuestra opinión, una aplicación incoherente de “las exenciones de la política sobre lenguaje que incita al odio para la expresión de grupos marginados”. Para abordar esta cuestión, recomendamos que Meta tradujera al árabe estándar moderno las “normas internas de implementación” y las “preguntas conocidas” que da a sus moderadores.

Uso gubernamental de las plataformas de Meta



Música drill británica

ANULADA

En enero de 2022, una cuenta de Instagram descrita como promotora de música británica publicó un clip de un video musical recién lanzado de un tema de “música drill británica” llamado “Secrets Not Safe” del rapero Chinx (OS). La música drill británica es un subgénero localizado y popular del rap, famoso entre los jóvenes negros urbanos del Reino Unido. Poco después, la Policía Metropolitana, un organismo policial con responsabilidades en el Gran Londres, envió un correo electrónico a Meta en el que solicitaba a la empresa que revisara todo el contenido asociado a “Secrets Not Safe”. La policía envió más información relacionada con la prevalencia de la violencia de bandas en Londres y expresó su preocupación por que el tema pudiera aumentar el riesgo de represalias violentas por parte de las bandas.

En respuesta a esta solicitud, un equipo de especialistas de Meta, en función del contexto proporcionado por la policía, determinó que la publicación constituía una “amenaza velada”. La empresa basó esa conclusión en una referencia a un tiroteo ocurrido en 2017, lo que, en su opinión, aumentaba el potencial del tema para incitar a la violencia. Meta no solo eliminó el contenido de la cuenta por infringir su política sobre violencia e incitación, sino que también eliminó de otras cuentas más de 50 contenidos que incluían el tema, incluso de la del artista. Posteriormente, los sistemas automatizados de Meta eliminaron contenido que incluía el tema otras 112 veces.

Después de que Meta remitió el caso al Consejo, pedimos revisar la publicación original de Chinx (OS). Meta respondió que las medidas que había tomado para eliminar el video y la canción de la plataforma habían provocado la eliminación de la cuenta del artista. A la hora de considerar si las acciones de Meta se alineaban con sus normas, valores y responsabilidades en materia de derechos humanos, el contexto de la publicación fue clave:

“Los artistas [de rap] suelen hablar con detalles minuciosos de los conflictos violentos que ocurren actualmente en las calles, con una narrativa en primera persona e imágenes y letras que describen o representan actos violentos. Las posibles declaraciones de violencia y bravuconería performativa se consideran parte del género (una forma de expresión artística donde la realidad y la ficción pueden confundirse).

A través de estas declaraciones, los artistas compiten por relevancia y popularidad. Si la música drill genera violencia en la vida real o no es un asunto controversial, particularmente en lo que respecta a la confiabilidad de las declaraciones probatorias que se hacen en los debates”.

Al evaluar si la decisión de Meta se alineaba con sus reglas, valores y responsabilidades en materia de derechos humanos, citamos la falta de pruebas que respaldaran la idea de que el contenido constituía una “amenaza creíble”. Concluimos además que, “ante la falta de dichas pruebas, Meta debería haber dado más peso a la naturaleza artística del contenido”. En nuestra decisión, señalamos que, “si bien las autoridades policiales en ocasiones

pueden proporcionar contexto y experiencia, no todo el contenido que las autoridades policiales preferirían eliminar debe eliminarse”.

Al evaluar dichas solicitudes en el futuro, instamos a Meta a adoptar medidas tangibles para “evaluar estas solicitudes de forma independiente, particularmente cuando se relacionan con expresiones artísticas de personas pertenecientes a minorías o grupos marginados, cuyo contenido corre un riesgo crítico de sesgo cultural”.

Nuestra decisión también resaltó la falta de transparencia sobre los procesos que utiliza Meta para responder a las solicitudes de los organismos gubernamentales y policiales. Citamos información que recibimos en respuesta a una solicitud de libertad de la información a la Policía Metropolitana, que revelaba que las 286 solicitudes que la policía había presentado a empresas de redes sociales y servicios de streaming para revisar o eliminar contenido musical en el año hasta mayo de 2022 se referían a música drill británica. Sin embargo, el 4 de enero de 2023, la Policía Metropolitana se puso en contacto con nosotros para decirnos que había identificado errores en su respuesta y los había corregido. En realidad, la Policía Metropolitana había presentado 992, no 286, solicitudes de este tipo en el año hasta mayo de 2022, todas ellas relacionadas con música drill.

Trato justo hacia los usuarios



Opinión consultiva sobre políticas acerca del uso compartido de información de residencia privada

En febrero de 2022, publicamos nuestra primera opinión consultiva sobre políticas (PAO). El Consejo utiliza las opiniones consultivas sobre políticas para revisar las políticas de Meta y formular recomendaciones sobre cómo mejorarlas. En este caso, Meta pidió al Consejo una evaluación detallada de los cambios que podría introducir en la norma comunitaria de Facebook sobre infracciones de la privacidad, la cual prohíbe compartir “información de identificación personal sobre uno mismo u otras personas”, salvo en condiciones específicas. Según los términos de esta política, la “información de identificación personal”, incluidas las direcciones residenciales, no puede compartirse, salvo en los casos en que se “comparta o solicite para promover causas benéficas, encontrar personas, animales u objetos perdidos, o ponerse en contacto con proveedores de servicios comerciales”.

Para proteger la privacidad de los usuarios, la política prohíbe compartir “imágenes donde aparezca la fachada de residencias privadas”, especialmente si la residencia es una vivienda unifamiliar, puede identificarse el número de vivienda en la imagen o el texto, el residente “se opone a que se exhiba su residencia privada” o si “hay contexto que indica que se organizarán protestas contra el residente”. La única excepción es si la residencia funciona como embajada del Gobierno, en cuyo caso puede ser foco de protestas.

Un problema subyacente que Meta nos pidió que abordáramos es el hecho de que tener acceso a esa información puede ser útil para periodistas y activistas cívicos. Por el contrario, exponer dicha información sin el consentimiento de los residentes podría “crear un riesgo para la seguridad de los residentes e infringir la privacidad de las personas”. Entre los posibles daños en el mundo real para la seguridad y la privacidad de los usuarios, Meta citó el “doxing”, término inglés que se refiere a la divulgación no autorizada de

documentos (de forma abreviada, “dox”) con la intención de revelar información personal a personas que podrían hacer un uso indebido de ella. Como señaló Meta, el doxing tiene consecuencias negativas en el mundo real, como el acoso, el hostigamiento, la violencia y la muerte.

La política contiene excepciones en virtud de las cuales la información de residencia privada puede publicarse; por ejemplo, cuando ya está “disponible públicamente a través de la cobertura de noticias, archivos judiciales, comunicados de prensa u otras fuentes”. En su orientación interna para los revisores de contenido, Meta afirma que la información publicada anteriormente por “al menos cinco medios de noticias” ya no se considera privada en virtud de la política.

Estamos de acuerdo con la afirmación de Meta de que las infracciones del derecho a la privacidad residencial pueden provocar daños graves en el mundo real y que dichos daños afectan de manera desproporcionada a las mujeres, los niños y las personas de la comunidad LGBTQIA+. También coincidimos con Meta en que, mientras que acceder físicamente a los registros públicos y a otras fuentes de información “disponible públicamente” requiere recursos y esfuerzo, acceder a esa información es mucho más fácil en las plataformas digitales, ya que se comparte con mayor facilidad y a mayor escala. En vista de la gravedad e intensidad de tales daños, aconsejamos a Meta que limitara las circunstancias en las que se permite compartir información “disponible públicamente”.



Opinión consultiva sobre políticas acerca del programa de verificación de Meta

En octubre de 2021, a raíz de las divulgaciones que realizó el *Wall Street Journal* sobre el programa de verificación de Meta, aceptamos una solicitud de la empresa para revisar el programa y formular recomendaciones sobre cómo mejorarlo. En el contexto de esa solicitud, Meta indicó que hacía unos 100 millones de intentos de aplicación de contenido todos los días. A ese ritmo, aunque el 99 % de sus decisiones de moderación fueran correctas, la empresa aún cometería alrededor de un millón de errores al día.

El enorme volumen y la complejidad del contenido publicado en Facebook e Instagram plantean verdaderos retos para la creación de sistemas que cumplan los compromisos de Meta en materia de derechos humanos. Aunque reconocemos esto, observamos que, tal como está estructurada actualmente, la verificación no cumplía los compromisos de la empresa en materia de derechos humanos. Durante años, la verificación permitió que el contenido de un grupo selecto de políticos, socios comerciales y celebridades, entre otros, continuara en Facebook e Instagram durante varios días, aunque dicho contenido se hubiera eliminado rápidamente en otros casos. Meta nos informó que, en promedio, puede tardar más de cinco días en tomar una decisión relacionada con el contenido de usuarios en listas de verificación.

Meta, por su parte, sostuvo que el objetivo original del programa había sido “promover sus compromisos en materia de derechos humanos”. Tras un examen más detallado, concluimos que el programa parecía “orientado más bien a satisfacer intereses comerciales”. Uno de los defectos más destacados que surgió a partir de nuestra revisión,

y que instamos a la empresa a abordar, fue que, al conceder a determinados usuarios una mayor protección frente a la eliminación de contenido que a otros, se amplificaba la preocupación generalizada de que la empresa favorecía a celebridades y líderes políticos y empresariales sobre millones de personas.

Si se detecta que una publicación de un usuario incluido en las listas de verificación de Meta infringe las reglas de la empresa, permanece en la plataforma en espera de otra revisión. Entonces, Meta aplica todas sus políticas a la publicación, incluidas las excepciones y las disposiciones específicas del contexto, lo que probablemente aumenta sus probabilidades de permanecer en la plataforma. En cambio, los usuarios comunes tienen muchas menos probabilidades de que su contenido llegue a revisores que puedan aplicar todas las reglas de Meta.

Este trato desigual se vio amplificado por la falta de transparencia en torno a los criterios utilizados para elaborar la lista de entidades cuyos contenidos se benefician de una protección adicional. Aunque Meta tiene criterios claros para agregar socios comerciales y líderes gubernamentales a las listas de verificación, los grupos cuyas publicaciones son importantes desde el punto de vista de los derechos humanos, como periodistas o líderes de la sociedad civil, tienen vías menos claras para acceder al programa.

Esta falta de transparencia se extiende al hecho de que Meta no revela los parámetros básicos que utiliza para medir la eficacia del programa. Meta no aportó pruebas de que la empresa compare la precisión de las decisiones que toma en virtud de la verificación con las que toman sus mecanismos normales de control de calidad. Se extiende al hecho de que Meta sigue sin informar a los usuarios si figuran en listas de verificación, no revela los procedimientos de creación y auditoría de esas listas y no revela los casos en que “entidades que publican contenido que infringe las normas reiteradamente siguen formando parte de las listas debido a su perfil”.

Formulamos 32 recomendaciones para abordar estos problemas. En ellas se instaba a Meta a dar prioridad en sus listas de verificación a las expresiones importantes para los derechos humanos, incluidas las expresiones de “especial importancia pública”. También pedimos a Meta que aumentara radicalmente la transparencia sobre el programa y cómo funciona, e hicimos propuestas para reducir el daño provocado por el contenido infractor que se deja durante la “revisión mejorada”. Meta respondió a nuestras recomendaciones en marzo de 2023 y se comprometió a mejorar la transparencia y a ampliar la protección para quienes corren un riesgo especial de aplicación excesiva, como los periodistas y los defensores de los derechos humanos.



Dudas sobre cómo pedir Adderall®

ANULADA

En junio de 2021, un usuario de Facebook en Estados Unidos, quien dijo tener trastorno por déficit de atención con hiperactividad (TDAH), preguntó cómo hablar con un médico acerca de un medicamento en una publicación de un grupo privado. El usuario indicó que le habían recetado Xanax, pero que Adderall le había funcionado bien en el pasado. El usuario expresó su preocupación de que, si pedía a su médico una receta diferente, podría ser visto como alguien “que busca drogas”.

En agosto de 2021, Meta eliminó la publicación citando una supuesta infracción de la norma comunitaria de Facebook sobre bienes y servicios restringidos. Después de que seleccionamos el caso, Meta marcó la eliminación como “error del sistema de cumplimiento de normas” y restauró la publicación.

En nuestra decisión, concluimos que la decisión inicial de Meta de eliminar la publicación era incorrecta. La norma comunitaria de Facebook sobre bienes y servicios restringidos no prohíbe a los usuarios pedir consejo sobre drogas farmacéuticas específicas cuando se trata de “afecciones”. También concluimos que las definiciones públicas de las sustancias en la norma son poco claras para los usuarios porque, incluso según sus definiciones internas, el Adderall y el Xanax podrían o no entrar en la categoría de drogas “no médicas” o farmacéuticas en función de las circunstancias. En nuestras recomendaciones, instamos a la empresa a “revisar las apelaciones de los usuarios de manera oportuna cuando las medidas de aplicación que se implementen para el contenido provoquen penalizaciones a la cuenta”.

Aplicación automatizada de políticas y selección del contenido



Caricatura sobre la policía colombiana

ANULADA

En septiembre de 2020, un usuario de Facebook en Colombia publicó una imagen de una caricatura parecida al escudo oficial de la Policía Nacional de Colombia. Mostraba a tres figuras con uniforme de policía que sostenían porras sobre sus cabezas, aparentemente pateando y golpeando a otra figura tendida en el suelo con sangre bajo la cabeza. En español, el texto del escudo decía, según la traducción de Meta: “República de Colombia - Policía Nacional - Bolillo y Pata”.

Dieciséis meses después de que se publicara el contenido, Meta lo eliminó tras cotejar la imagen con otra similar en uno de sus bancos del Servicio de coincidencias. Estos bancos pueden identificar y eliminar automáticamente las imágenes que, durante las revisiones manuales, se considera que infringen las reglas de la empresa. Después de seleccionar el caso, Meta determinó que la publicación no infringía sus reglas y la restauró. La empresa también restauró otro contenido en el que aparecía la caricatura y reconoció que sus bancos del Servicio de coincidencias lo habían eliminado incorrectamente.

“Las decisiones sobre el contenido y el tratamiento algorítmico de este no pueden separarse”.

Catalina Botero-Marino
COPRESIDENTA DEL CONSEJO
ASESOR DE CONTENIDO



Este caso resaltó cómo estos bancos pueden aumentar el impacto de las decisiones incorrectas a la hora de incluir contenido en el banco. A pesar de que 215 usuarios apelaron estas eliminaciones, y el 98 % obtuvo buenos resultados, Meta no eliminó la caricatura de su banco ni restauró el contenido comparable hasta que tomamos el caso. Instamos a Meta a “mejorar de inmediato sus procedimientos para eliminar con rapidez contenido no infractor de estos bancos”.



Video tras el ataque en una iglesia de Nigeria

ANULADA

El 5 de junio de 2022, terroristas atacaron una iglesia católica en el suroeste de Nigeria. Horas más tarde, un usuario de Instagram en Nigeria publicó un video en el que se veían cuerpos inmóviles y ensangrentados en el suelo de la iglesia, algunos con la cara visible. Después de que uno de los bancos del Servicio de coincidencias de Meta identificara que el contenido se parecía a un video que una revisión manual había considerado como infractor, un “clasificador” automatizado de contenido decidió que podía permanecer publicado, con una pantalla de advertencia de “contenido perturbador”.

Varios días después, mientras un segundo banco de coincidencias de contenido multimedia revisaba el contenido, el usuario agregó un texto en inglés en el que describió el ataque como “triste”. El usuario también agregó hashtags, algunos de los cuales hacían referencia al videojuego de acción real de “airsoft”, así como otros utilizados habitualmente para vender armas de fuego a coleccionistas. Meta determinó que, si bien el video en sí no infringía las normas, los hashtags sí lo hacían, ya que podían interpretarse como que “exaltan la violencia y minimizan el sufrimiento de las víctimas”. La mayoría del Consejo no estuvo de acuerdo con la decisión de Meta y concluyó que la restauración original del contenido con una pantalla de advertencia había sido correcta. La mayoría basó su conclusión en la determinación de que los hashtags no deben clasificarse como sádicos “solo porque están asociados con usuarios de armas de fuego”. Una minoría no estuvo de acuerdo con esta evaluación y señaló que el uso de hashtags relacionados con tiroteos podría interpretarse como “sádico y podría traumatizar a los sobrevivientes o a los familiares de las víctimas”. Señalando esta ambigüedad, recomendamos a Meta que revisara el lenguaje de la política pública sobre contenido violento y gráfico.

Género



Video sobre acoso sexual en la India

CONFIRMADA

En marzo de 2022, una cuenta de Instagram descrita como una “plataforma para las perspectivas dalit” publicó un video de la India en el que se veía a un grupo de hombres agrediendo a una persona identificada como “mujer tribal”. Los miembros de la comunidad “dalit”, antes conocidos como “intocables”, sufren la opresión del sistema de castas del país. A pesar de la ausencia de desnudos y de que la mujer en cuestión no podía identificarse, Meta eliminó el contenido por infringir la política de explotación sexual de adultos. Más tarde, los equipos internos de Meta revirtieron esa decisión y restauraron el contenido tras aplicar una “concesión de interés periodístico” y una pantalla de advertencia, en vista de su intención de concientizar sobre la discriminación que sufren desde hace mucho tiempo las “mujeres tribales” en la India.

Si bien consideramos correcta la decisión de Meta de restaurar el contenido tras aplicar una concesión de interés periodístico y una pantalla de advertencia, reiteramos nuestra preocupación por el hecho de que aplicar la concesión de interés periodístico a un contenido que, por lo demás, es infractor no es la forma correcta de “tratar estos casos a escala”.

“La concesión de interés periodístico es imprecisa, deja un amplio margen de discreción a quien la aplique, y no se puede garantizar su aplicación de manera coherente a escala. Tampoco incluye criterios claros para evaluar el daño potencial generado por el contenido que infringe la política sobre explotación sexual de adultos. El Consejo determina que las responsabilidades de Meta en materia de derechos humanos hacen que resulte necesario proporcionar normas más claras y procesos de aplicación más eficaces para este tipo de casos”.

INFORME SOBRE LA PUNTUALIDAD DE LA IMPLEMENTACIÓN DE NUESTRAS DECISIONES Y RESPUESTA A ESTAS POR PARTE DE META

- Según nuestros estatutos, Meta debe implementar nuestras decisiones en un plazo de siete días a partir de su publicación.
- Para las 12 decisiones que publicamos en 2022, Meta restauró o eliminó el contenido en este plazo de siete días, salvo en los casos en que el contenido ya se había restaurado.
- A través de nuestro Comité de implementación, conformado actualmente por cinco miembros del Consejo, seguimos instando a Meta a ofrecer una mayor transparencia sobre cómo identifica y adopta medidas de aplicación respecto del contenido que es idéntico al que aparece en nuestras decisiones y que se presenta en un contexto paralelo. Esto garantizaría que nuestras decisiones se aborden fuera del caso concreto y se generalicen a un contenido pertinente en contextos similares.

Aplicación de normas internacionales sobre derechos humanos a la moderación de contenido: artículo 19

Un tema que define el trabajo del Consejo es nuestra convicción de que Meta tomará decisiones de moderación de contenido de una manera más justa y fundamentada si las basa en las normas internacionales sobre derechos humanos con las que se ha comprometido. Para ello, nuestra acta constitutiva establece que “debemos prestar particular atención al impacto de eliminar contenido teniendo en cuenta las normas de derechos humanos que protegen la libertad de expresión”. Estas normas incluyen el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos (ICCPR), que establece que, si bien “toda persona tiene derecho a la libertad de expresión, el ejercicio de [ese] derecho puede estar sujeto a ciertas restricciones, pero solo según lo disponga la ley y sea necesario”. En el artículo 19, se proporciona una prueba en tres partes para evaluar las restricciones a la expresión:



1. ¿Cumple la restricción el principio de **legalidad**?

Analizamos si las reglas en las que se basó Meta para tomar su decisión son accesibles y lo suficientemente claras para que los usuarios las entiendan y las sigan. Es importante que las reglas sean claras para que quienes deben aplicarlas puedan tomar decisiones justas y coherentes.

2. ¿Tiene la restricción propuesta un **objetivo legítimo**?

Analizamos si la regla en la que se basa una decisión sigue un objetivo compatible con los derechos.

3. ¿La restricción propuesta fue **necesaria y proporcional**?

¿Fue la eliminación del contenido el medio menos intrusivo para lograr el objetivo, y fue la restricción proporcional a los intereses protegidos?



“Las empresas deben respetar el derecho a la libertad de expresión, que incluye difundir y recibir ideas de todo tipo, independientemente de las fronteras”.

Evelyn Aswad
COPRESIDENTA DEL CONSEJO
ASESOR DE CONTENIDO



Normas internacionales sobre derechos humanos en el proceso de toma de decisiones del Consejo

En 2021, Meta publicó una Política corporativa de derechos humanos en la que describió sus compromisos en materia de derechos humanos tomando como base los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas (UNGP). En este marco de 2011, se proporcionó una nueva orientación a las empresas privadas sobre sus responsabilidades para respetar los derechos humanos.

En su primer informe sobre derechos humanos,¹ publicado en 2022, Meta formuló su decisión de crear el Consejo como parte de sus esfuerzos para proporcionar acceso a soluciones para los impactos en los derechos humanos. Los UNGP piden a las empresas que proporcionen “acceso a soluciones”, por ejemplo, estableciendo “mecanismos de reclamación eficaces a nivel operativo para las personas y las comunidades que puedan verse afectadas [por sus operaciones]”.

Como mecanismo de reclamación a nivel operativo, el Consejo trata de aplicar los criterios de eficacia establecidos en los UNGP, por ejemplo, siendo legítimo, accesible, equitativo, transparente, compatible con los derechos y una fuente de aprendizaje continuo. Seleccionamos casos, emitimos decisiones y formulamos recomendaciones para impulsar en Meta el respeto de los derechos humanos de todas las personas. En cada decisión, brindamos un análisis detallado de las implicaciones y preocupaciones en materia de derechos humanos que motivan el caso.

PRESENTACIONES DEL CONSEJO ANTE LAS NACIONES UNIDAS

Tomando como base nuestro análisis de las responsabilidades de Meta en materia de derechos humanos en casos y opiniones consultivas sobre políticas, realizamos presentaciones extensas a dos órganos de las Naciones Unidas en 2022 en las que describimos la evolución de nuestro enfoque para aplicar los principios de derechos humanos a las políticas y decisiones sobre el contenido de las plataformas digitales.

En nuestra presentación de febrero ante la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos acerca de “la aplicación práctica de los UNGP a las actividades de las empresas tecnológicas”² y en nuestra presentación de julio ante la Relatoría Especial de las Naciones Unidas sobre la libertad de opinión y de expresión³, nos centramos en el par de temas referentes a la “responsabilidad” y la “solución”, tal como se establecen en los UNGP.

“Recordamos constantemente a Meta que los principios de los derechos humanos deben ser la base de sus normas comunitarias”.

Nighat Dad
MIEMBRO DEL CONSEJO
ASESOR DE CONTENIDO



1 Meta Human Rights Report: Insights and Actions 2020 – 2021 (July 2022),” https://about.fb.com/wp-content/uploads/2022/07/Meta_Human-Rights-Report-July-2022.pdf

2 “Operationalizing the UN Guiding Principles on Business and Human Rights Submission to the Office of the High Commissioner for Human Rights, United Nations on the practical application of the UNGPs to the activities of technology companies”, Consejo asesor de contenido, febrero de 2022.

3 “Submission to the Special Rapporteur on freedom of opinion and expression: Challenges in Times of Conflicts and Disturbances,” Consejo asesor de contenido, julio de 2022.

Ambas presentaciones destacaron las serias implicaciones en materia de derechos humanos que tienen las decisiones de las redes sociales sobre el contenido durante los conflictos. Este enfoque refleja la importancia que atribuimos a los UNGP para que las empresas “aumenten sus esfuerzos para minimizar los riesgos en materia de derechos humanos donde las consecuencias sean ‘más graves’”. En consonancia con los UNGP, para cada caso asignado a un panel, realizamos una detallada Evaluación del impacto en los derechos humanos (EIDH), que puede incluir un “análisis de sensibilidad del conflicto” para garantizar que el propio Consejo cumple con sus responsabilidades de diligencia debida. Además, nuestro Equipo de selección de casos cuenta con procesos para alertar a Meta sobre el contenido que se “deja” y puede necesitar una “atención expeditiva” para mitigar el riesgo de daños graves.

ESTUDIOS DE CASO DE AFGANISTÁN, ETIOPÍA Y SUDÁN

En la presentación ante la Relatoría Especial de las Naciones Unidas, citamos tres casos ilustrativos de 2022: “Video gráfico de Sudán”, “Mención de los talibanes en los informes de noticias” y “Consejo de comunicaciones de Tigray”. Todos reflejaban la falta de avances de Meta en el desarrollo de un “sistema transparente y fundamentado para moderar el contenido en zonas de conflicto”. Sostuvimos que la formulación de un protocolo eficaz de respuesta a las crisis mejoraría significativamente la capacidad de la empresa para mantener sus responsabilidades en materia de derechos humanos durante situaciones de crisis y conflicto.

En nuestra decisión sobre el “**video gráfico de Sudán**”, subrayamos el contexto de crisis y conflicto en el momento en que se publicó el contenido: “Las fuerzas de seguridad apuntaron a periodistas y activistas, y registraron sus hogares y oficinas. Los periodistas [fueron] atacados, detenidos y arrestados. Tras la toma del poder por parte de los militares de los medios de comunicación estatales y las medidas estrictas contra



los periódicos y las emisoras de Sudán, las redes sociales se convirtieron en una fuente fundamental de información y en un lugar donde documentar la violencia que ejercen los militares. Los militares desactivaron Internet y detuvieron a líderes civiles”.

Al aplicar la prueba en tres partes del artículo 19, cuestionamos la política de Facebook sobre contenido violento y gráfico por no aclarar cómo permite a los usuarios “compartir contenido gráfico para generar conciencia sobre los abusos o documentarlos”. Aunque Meta aplicó correctamente su “concesión de interés periodístico” como fundamento principal para restaurar el contenido en cuestión, señalamos que la regla no define con claridad el término pertinente. En cuanto a la solución de colocar una pantalla de advertencia sobre el contenido, sostuvimos que este método constituía una “restricción necesaria y proporcional a la libertad de expresión”, que “protege de manera adecuada la dignidad de la persona que se muestra y su familia”.

En el segundo caso, **“Mención de los talibanes en los informes de noticias”**, acerca de la eliminación de una publicación que difundía el anuncio de un portavoz oficial de los talibanes en Afganistán sobre la inminente apertura de escuelas para niñas, consideramos que la decisión inicial de Meta de eliminar el contenido infringía la libertad de expresión de los usuarios. Esto se debió a que negó a los usuarios de Facebook en Afganistán su derecho a “acceder a información sobre los acontecimientos de interés público, especialmente cuando un grupo peligroso designado expulsó por la fuerza al Gobierno reconocido”. Agregamos que las restricciones a la libertad de los medios de comunicación impuestas por el régimen talibán reforzaron “aún más el rol de la cobertura periodística internacional” porque “la información era esencial para las personas preocupadas por el derecho igualitario de las niñas y las mujeres a obtener educación”.



Si se va a reprimir la libertad de expresión, debe quedar claro por qué”.

Julie Owono
MIEMBRO DEL CONSEJO
ASESOR DE CONTENIDO



En el tercer caso, **“Consejo de comunicaciones de Tigray”**, acerca de la eliminación de contenido que incluía una petición de una fuente oficial del Gobierno provincial de Tigray para que las fuerzas gubernamentales se enfrentaran a las fuerzas del Gobierno central, se concluyó que la decisión de Meta de eliminar la publicación fue coherente con sus responsabilidades en materia de derechos humanos. Para respaldar esa decisión, aplicamos la prueba de los seis factores del Plan de Acción de Rabat, que proporciona orientación específica sobre cómo proteger la libertad de expresión y, al mismo tiempo, proteger a las personas contra la incitación a la discriminación, la hostilidad o la violencia.

Los seis factores de Rabat se evaluaron del siguiente modo: (1) *Contexto*: El contenido se publicó en el contexto de una guerra civil en curso y cada vez más grave. (2) *Orador*: El orador fue un ministerio del Gobierno regional afiliado a una de las partes del conflicto. (3) *Intención*: Un pedido explícito de matar a los soldados que no se rindieran. (4) *Contenido*: Podría interpretarse que la publicación promovía apuntar a combatientes y líderes políticos, independientemente de su participación en las hostilidades. (5) *Alcance de la difusión*: El contenido se publicó en la página pública de un organismo vinculado a una de las partes del conflicto, con unos 260.000 seguidores, y permaneció en la plataforma durante 2 días antes de su eliminación. (6) *Probabilidad e inminencia*: El contenido se publicó mientras las fuerzas de Tigray avanzaban más allá de Tigray y el primer ministro declaraba el estado de emergencia en todo el país y pedía a los civiles que tomaran las armas y lucharan.

También empleamos la prueba de Rabat para respaldar nuestra decisión en el caso sobre los “dibujos animados de Knin” de 2022, acerca de la publicación de un dibujo animado que caracterizaba como ratas a las personas de etnia serbia en una ciudad de Croacia. (1) *Contexto:* Una región que ha sufrido recientemente conflictos étnicos y discriminación contra las minorías étnicas. (2) *Orador:* Un portal de noticias croata conocido por sus sentimientos antiserbios. (3) *Intención:* Incitar al odio étnico. (4) *Contenido:* El formato de video de dibujos animados puede ser particularmente dañino porque es atractivo. (5) *Alcance de la difusión:* El contenido se vio más de 380.000 veces, se compartió más de 540 veces, recibió más de 2.400 reacciones y tuvo más de 1.200 comentarios. (6) *Probabilidad e inminencia:* El Consejo no consideró que la publicación pudiera provocar un daño inminente. No obstante, señalamos la decisión de que Meta “puede, de forma legítima, eliminar de Facebook las publicaciones que fomenten la violencia de un modo menos inmediato”.



Participación y comentarios públicos

Como Consejo, sabemos que solo podemos encontrar soluciones duraderas a los retos de la moderación de contenido si escuchamos a otras organizaciones y trabajamos con ellas. Con este objetivo, los miembros del Consejo y del personal participan en actos y debates públicos, en los que explican la labor del Consejo y hablan sobre por qué es importante para los usuarios obtener un enfoque global y fundamentado de la moderación de contenido. Como parte de nuestro proceso de decisiones, las personas y las organizaciones también pueden enviar comentarios públicos. Estos han dado voz a las personas en nuestras decisiones y han aportado experiencia en idiomas, cultura, política y derechos humanos. En múltiples ocasiones, los comentarios públicos también han influido en nuestras decisiones y recomendaciones a Meta.

||| DIVULGACIÓN Y PARTICIPACIÓN

En 2022, más de 1.000 partes interesadas asistieron a las mesas redondas que organizamos sobre cuestiones planteadas en casos concretos. Entre los temas tratados, figuró nuestra opinión consultiva sobre políticas acerca de la verificación y las políticas de Meta en materia de información errónea sobre el COVID. También distribuimos el boletín “Across the Board” para informar mejor al público sobre las actividades del Consejo. En 2023, continuaremos esta labor fundamental convocando al menos una mesa redonda al mes en lugares de todo el mundo con enfoque en nuestras siete prioridades estratégicas. Al centrar nuestro trabajo de



participación en torno a estas siete prioridades, también esperamos beneficiarnos de nuestras redes y de las comunidades académicas y de defensa más extensas que se especializan en estas áreas.

COMENTARIOS PÚBLICOS

Para enriquecer nuestras decisiones y opiniones consultivas sobre políticas, consideramos de forma cuidadosa los comentarios públicos que envían personas y organizaciones. En 2022, recibimos más de 600 comentarios públicos, con un aumento en el número de comentarios públicos enviados en la segunda mitad del año.

La decisión del **“Consejo de comunicaciones de Tigray”** se basó en las valiosas percepciones de expertos en moderación de contenido en zonas de conflicto. Una de las conclusiones, que “el enfoque actual de Meta respecto a la moderación de contenido en zonas de conflicto [puede] generar una apariencia de incoherencia”, se basó en los comentarios enviados por varios expertos, incluido el Dr. Samson Esayas, profesor asociado de la Escuela de Negocios de Noruega BI, quien señaló una disparidad patente entre las “medidas rápidas” de Meta para moderar contenido en el contexto europeo del conflicto entre Rusia y Ucrania, y su aparente “trato diferenciado entre este conflicto y los conflictos en otras regiones, particularmente Etiopía y Birmania”.

Nuestra decisión sobre **“música drill británica”** se basó en los conocimientos especializados de varias organizaciones. La Digital Rights Foundation argumentó que emplear un análisis lírico línea por línea de la canción eliminada para determinar las pruebas de infracciones pasadas o de riesgo de daños futuros es notoriamente inexacto y que verificar afirmaciones presuntamente factuales dentro de las letras de la música drill es difícil. La Electronic Frontier Foundation también criticó el control de la música drill legal por parte de las autoridades policiales. También se recibió un comentario público del Servicio de Policía Metropolitana, el primero de un organismo gubernamental. Aunque la Policía Metropolitana no dio su consentimiento para publicar el comentario, indicó que podría darlo más adelante.

Nuestra opinión consultiva sobre políticas acerca del **programa de verificación de Meta** también provocó una amplia variedad de comentarios. Organizaciones como Center for Democracy and Technology, el Instituto de Tecnología y Sociedad de Río de Janeiro, PEN America y Mnemonic compartieron su experiencia con el Consejo. Los políticos demócratas y republicanos también compartieron sus preocupaciones sobre la verificación mediante el envío de comentarios públicos.

Los casos que recibieron el mayor número de comentarios en 2022 se publicaron a principios de 2023. Nuestro caso sobre el **“eslogan de protesta contra Irán”** suscitó 162 comentarios públicos. Los comentarios públicos confirmaron que el término “marg bar Khamenei” (que literalmente se traduce como “muerte al [líder supremo de Irán, Ali] Khamenei”) se utilizó ampliamente durante las recientes protestas en Irán. También expresaron su preocupación porque Meta había eliminado incorrectamente contenido en idioma farsi durante estas protestas.

En nuestros casos de “identidad de género y desnudos”, muchos de los más de 100 comentarios que recibimos provenían de personas que se identifican como mujeres trans, no binarias o cisgénero. Muchas de las personas que enviaron los comentarios señalaron que, al igual que los usuarios en estos casos, también habían experimentado la eliminación incorrecta de sus publicaciones. En los comentarios, se expresaba confusión

“Hemos analizado información valiosa en los comentarios públicos para poder tomar mejores decisiones”.

Ronaldo Lemos
MIEMBRO DEL CONSEJO
ASESOR DE CONTENIDO



acerca de por qué Meta eliminaba las publicaciones que incluían enlaces a sitios web de recaudación de fondos. En otro comentario, ACON, una ONG de Australia dedicada a la educación sobre el VIH, escribió que se habían eliminado contenidos que promovían mensajes de prevención del VIH por solicitudes sexuales. Esto lo repitió Joanna Williams, una investigadora que descubrió que 9 de las 12 organizaciones de salud sexual a las que entrevistó afirmaron verse afectadas de forma negativa por la moderación de Meta en esta área. Un comentario del centro de investigación InternetLab expresó su preocupación por la presunta sexualización de los cuerpos de las mujeres, trans y no binarias, cuando no se aplica una suposición comparable a los hombres cisgénero. El impacto desproporcionado de la eliminación de contenido con cuerpos de mujeres también se señaló en un comentario de la Dra. Zahra Stardust. Por último, los comentarios de la Gay & Lesbian Alliance Against Defamation (GLAAD) y la Human Rights Campaign Foundation plantearon la preocupación de que el contenido de usuarios de grupos marginados corre un mayor riesgo de sufrir denuncias repetidas o malintencionadas, en las que los usuarios reportan contenido no infractor para acosar a la persona que lo ha publicado.

Nuestra **opinión consultiva sobre políticas acerca de la eliminación de información errónea sobre el COVID-19** recibió 181 comentarios públicos. Este fue el mayor número de comentarios que recibimos para una sola decisión u opinión en 2022. Un comentario de American Civil Liberties Union planteaba la preocupación de que la dificultad para distinguir, a escala, entre realidad y ficción, y entre opinión, experiencia y afirmación de hechos significaba que Meta reprimiría discursos que deberían estar permitidos.

Varios comentarios señalaron la responsabilidad de Meta de abordar los riesgos para la seguridad pública, dado su alcance y el rol de sus sistemas en la amplificación de la información errónea. Se expresó preocupación por la idoneidad de las etiquetas y las degradaciones al abordar el riesgo de daños. Por ejemplo, el comentario del vicepresidente sénior del Centro de Internet y Estudios Internacionales destacaba su preocupación de que las etiquetas eran insuficientes para abordar la información errónea que difundieron políticos y personas influyentes prominentes.



Cronología de las actividades de participación en 2022

En 2022, los miembros del Consejo, los fiduciarios y el personal de la Administración participaron en casi 100 eventos en todo el mundo. La cronología de actividades de participación que aparece a continuación ofrece algunos ejemplos de estos eventos.

MARZO

Un representante del Consejo habla acerca del valor de la supervisión independiente de contenido digital en el **Congreso Mundial de Móviles** en Barcelona.

MAYO

Un miembro del Consejo da un discurso en el **Foro de la Libertad de Oslo** acerca del rol potencial que tienen las redes sociales en el fortalecimiento de la democracia y la resistencia a la tiranía.

JUNIO

Miembros del Consejo y un fiduciario participaron en **RightsCon**, la principal cumbre del mundo sobre “derechos humanos en la era digital”. En los paneles se trataron temas como la libertad de expresión en línea en Irán y la gobernanza multipartes.

Miembros del Consejo se reúnen en persona por primera vez en California, para celebrar reuniones con **líderes sénior de Meta, fiduciarios del Consejo** y un grupo de partes interesadas.

SEPTIEMBRE

Un miembro del Consejo habla en la **Cumbre Anual de Concordia** que se celebra en paralelo con la **Asamblea General de las Naciones Unidas** en Nueva York.

OCTUBRE

Miembros del Consejo se unen a expertos del sector en Nueva York para hablar sobre el futuro de la moderación de contenido en la conferencia de Columbia Global Freedom of Expression

NOVIEMBRE

Un miembro del Consejo habla en la **Web Summit** anual en Lisboa.

En el congreso **Bread and Net** en **Beirut**, un miembro del Consejo participa en una mesa redonda a la que asistieron partes interesadas de Oriente Medio.

Próximos pasos

2023 Y MÁS ALLÁ

Evolucionar nuestro trabajo con Meta

En 2022, perfeccionamos nuestro trabajo con Meta, publicamos nuestras primeras opiniones consultivas sobre políticas y seguimos generando un impacto a través de nuestras decisiones y recomendaciones. En 2023, construiremos sobre estos fuertes cimientos. En respuesta a las opiniones de las partes interesadas, seguimos repitiendo nuestro enfoque y hemos fijado seis objetivos para nuestro trabajo con Meta.

Objetivos del Consejo asesor de contenido para 2023:

1. Publicar nuestras primeras decisiones sumarias
2. Emitir nuestras primeras decisiones expeditas
3. Alcanzar nuestro objetivo actualizado de miembros del Consejo para lograr la máxima eficacia
4. Reforzar el compromiso en torno a nuestras siete prioridades estratégicas
5. Proseguir los planes a largo plazo para ampliar el ámbito
6. Supervisar la forma en que Meta implementa nuestras recomendaciones y pedir a la empresa que proporcione pruebas de su implementación e impacto

1. PUBLICAR NUESTRAS PRIMERAS DECISIONES SUMARIAS.

De los 50 casos que seleccionamos en 2022, Meta revirtió de forma inmediata su decisión original en 32 de ellos y restauró o eliminó el contenido según fuera necesario. En total, Meta ha eliminado o restaurado más de 80 publicaciones desde 2020 debido a que el Consejo señaló estos errores; gracias a ello, los usuarios obtuvieron enmiendas y se revelaron percepciones fundamentales sobre cómo la empresa modera el contenido. En 2023, los miembros del Consejo en el Comité de selección de casos seleccionarán algunos de estos casos para revisarlos como decisiones sumarias. En ellas se establecerá por qué consideramos que el caso es significativo y señalaremos las recomendaciones que hemos hecho en situaciones similares en decisiones y opiniones consultivas sobre políticas anteriores. El Comité de selección de casos redactará las decisiones sumarias. Las aprobará y someterá a votación el Comité de selección de casos, no el Consejo en pleno; asimismo, no se tendrán en cuenta los comentarios públicos.

2. EMITIR NUESTRAS PRIMERAS DECISIONES EXPEDITAS.

En 2023, esperamos publicar nuestras primeras decisiones expeditas. Meta nos remitirá estos casos de forma expedita. Redactar y publicar una decisión en cuestión de días nos permitirá contribuir en temas de interés público y situaciones con impactos graves en los derechos humanos a medida que se produzcan. Un panel deliberará y redactará y aprobará una decisión por escrito, que se publicará en nuestro sitio web. Las decisiones expeditas sobre eliminar o dejar contenido serán vinculantes para Meta. Por restricciones de tiempo, estos casos no considerarán los comentarios públicos y se decidirán sobre la base de la información disponible en el momento de la deliberación.

3. ALCANZAR NUESTRO OBJETIVO ACTUALIZADO DE MIEMBROS DEL CONSEJO PARA LOGRAR LA MÁXIMA EFICACIA.

Aunque originalmente esperábamos que el Consejo alcanzara los 40 miembros, 3 años de operaciones nos han demostrado que, en la práctica, el número óptimo de miembros que permite una deliberación y una toma de decisiones oportunas, regulares y eficaces sería 26. Después de la renovación de la mayoría de nuestros actuales miembros del Consejo en abril de 2023, y una vez completados los procesos en curso para la selección de 2 nuevos miembros del Consejo, Meta se retirará del proceso de selección. A partir de ese momento, la selección de los miembros del Consejo estará a cargo de los miembros del Consejo y los fiduciarios, sin la participación de Meta. La sustitución de cualquier miembro del Consejo ha sido y seguirá siendo a discreción de los miembros del Consejo y los fiduciarios. Esperamos que el Consejo cuente con un conjunto completo de miembros para finales de 2023.

4. REFORZAR EL COMPROMISO EN TORNO A NUESTRAS SIETE PRIORIDADES ESTRATÉGICAS.

En 2022, elegimos siete prioridades estratégicas en función de un análisis de los casos presentados al Consejo y los problemas a los que se enfrentan los usuarios en todo el mundo. En 2023, estas prioridades guiarán los casos que Meta nos remita y aquellos que finalmente seleccionemos para revisar. Para todas nuestras prioridades, continuaremos trabajando con las partes interesadas para entender las políticas y las prácticas de aplicación que Meta necesita mejorar con mayor urgencia, así como qué tipo de casos podría brindar la oportunidad de abordarlas. Animamos a las partes interesadas especializadas en estas áreas a que nos contacten a través de nuestros procesos de comentarios públicos, mesas redondas y conversaciones individuales.

5. PROSEGUIR LOS PLANES A LARGO PLAZO PARA AMPLIAR EL ÁMBITO.

En 2022, pudimos tomar decisiones vinculantes para aplicar una pantalla de advertencia al dejar o restaurar contenidos que cumplieran los requisitos. En 2023, seguiremos dialogando con Meta sobre la ampliación de nuestro ámbito para incluir grupos y cuentas. Aunque prevemos que este trabajo comenzará en 2023, Meta nos ha comunicado que, por razones técnicas y operativas, es poco probable que los grupos y las cuentas entren en el ámbito antes de 2024. Como Consejo, también estamos interesados en explorar la ampliación del ámbito en otras áreas, incluida la amplificación y degradación de contenidos.

6. SUPERVISAR LA FORMA EN QUE META IMPLEMENTA NUESTRAS RECOMENDACIONES Y PEDIR A LA EMPRESA QUE PROPORCIONE PRUEBAS DE SU IMPLEMENTACIÓN E IMPACTO.

Continuaremos supervisando de cerca cómo Meta implementa nuestras recomendaciones y proporcionando actualizaciones en nuestros informes de transparencia trimestrales. También pediremos a la empresa que aporte pruebas de la implementación de las recomendaciones en sus políticas y productos, y que comparta parámetros sobre el impacto de estas recomendaciones en la experiencia de las personas que utilizan sus plataformas y se ven afectadas por ellas.

Compartir los beneficios de la supervisión independiente

En 2022, un número creciente de empresas solicitó experiencia externa en decisiones de moderación de contenido. Spotify creó un Comité asesor de seguridad y el sitio de microblogueo Koo creó un Consejo asesor. El antiguo responsable de Confianza y Seguridad de Twitter también pidió que se establecieran comités de moderación de contenido para las tiendas de aplicaciones de Google y Apple.

En esencia, la supervisión independiente consiste en que las empresas abran sus procesos internos e inviten a personas ajenas a revisar sus decisiones. Este tipo de cuestionamiento y escrutinio genera decisiones mejores y más sólidas, lo que ayuda a crear confianza con los usuarios a largo plazo: esta es una situación beneficiosa tanto para las comunidades como para las empresas.

Desde el principio, el Consejo asesor de contenido no se limitó a mejorar Facebook e Instagram. Esta también fue una oportunidad para experimentar con un nuevo enfoque independiente de moderación de contenido que podría adaptarse a otras plataformas y empresas tecnológicas. Como Consejo, hemos desarrollado una gran experiencia en los últimos tres años que podría ayudar a otras empresas a tomar mejores decisiones y brindar un mejor servicio a sus usuarios.



La libertad de expresión de miles de millones de personas en el planeta es demasiado importante para dejarla a cargo de una sola empresa”.

Alan Rusbridger
MIEMBRO DEL CONSEJO
ASESOR DE CONTENIDO



UNA EXPERIENCIA ÚNICA DE CREACIÓN DE UN CONSEJO ASESOR DE CONTENIDO

La creación y administración de un consejo asesor para la gobernanza de contenido es algo que nadie había intentado antes y que resultó ser más complicado de lo que parecía. Desde el primer día, nos esforzamos por superar un gran número de retos, que incluían desde la infraestructura técnica hasta la creación de una cultura de trabajo que equilibrara el rápido ritmo de la moderación de contenido con deliberaciones sensatas. Durante meses, trabajamos con Meta para crear un sistema de apelación independiente accesible para miles de millones de usuarios de todo el mundo. Establecimos un proceso de comentarios públicos para dar voz a las personas en nuestro proceso de toma de decisiones. También aprendimos a deliberar sobre casos sin respuestas fáciles teniendo en cuenta nuestras diferentes nacionalidades, orígenes y puntos de vista.

En nuestros tres años de funcionamiento, hemos aprendido y aplicado muchas lecciones acerca de cómo debe funcionar un consejo asesor para la gobernanza de contenido. En particular, identificamos cinco características que podrían ayudar a otras empresas tecnológicas a establecer este tipo de supervisión.

- ◉ **Independencia:** Cualquier órgano de supervisión debe estructurarse para permitir un juicio independiente. Para tener éxito en el escrutinio de las decisiones de contenido, debe estar libre de las restricciones comerciales, reputacionales o políticas de las plataformas. Como Consejo, no hemos dudado en anular las decisiones de Meta. Esta separación es fundamental para generar legitimidad con los usuarios y los grupos de la sociedad civil de todo el mundo.
- ◉ **Transparencia:** Para confiar en las decisiones de una empresa, las personas tienen que entender cómo se toman esas decisiones. Al preparar nuestras decisiones, incluimos una gran cantidad de información que previamente no era pública sobre las políticas y su aplicación. Para que se las considere como entidades responsables y genuinas en su compromiso de defender la libertad de expresión, las empresas deben explicar a los usuarios por qué se eliminaron sus publicaciones o se desactivaron sus cuentas. También deben adoptar aún más medidas para explicar por qué se toman tales decisiones y ser más transparentes cuando los Gobiernos o los agentes estatales les piden que eliminen contenidos.
- ◉ **Diversidad:** La mayoría de los usuarios de las plataformas de redes sociales se encuentran fuera de Estados Unidos y Europa. Muchas de las preocupaciones acerca de los impactos negativos de las redes sociales, así como de sus beneficios, se sienten con mayor intensidad en países situados fuera de Estados Unidos y Europa. Para poder confiar en la toma de decisiones de una empresa, las personas de estos países necesitan algo más que un simple “compromiso” con su parte del mundo. Necesitan sentirse escuchadas y representadas en la mesa de toma de decisiones. Por ello, los miembros del Consejo asesor de contenido son personas que han vivido en 27 países y que hablan 29 idiomas. La diversidad enriquece nuestras decisiones finales.
- ◉ **Derechos humanos:** En la actualidad, las empresas tecnológicas se enfrentan a un reto importante con respecto a la moderación de contenido: ¿qué reglas tienen que aplicarse a miles de millones de personas de diferentes nacionalidades, idiomas y culturas? Creemos que las normas internacionales sobre derechos humanos son una parte fundamental de la respuesta. Se aplican por igual a todos y proporcionan un marco coherente para considerar el derecho de un usuario a la libertad de expresión junto con otros derechos humanos, como el derecho a la vida o a la privacidad.
- ◉ **Asociación:** Finalmente, lo que se necesita es asociación. Los organismos de supervisión independiente solo tendrán un impacto duradero si las empresas les dan acceso a sus datos y procesos. Aunque a veces hay ocasiones en las que compartir información podría comprometer la privacidad de los usuarios o permitir a agentes perniciosos manipular las políticas, nuestro trabajo depende de datos e información que solo Meta puede proporcionar. Este tipo de transparencia también ayuda a crear legitimidad y confianza con los usuarios y la sociedad civil. Las empresas también tienen que mostrar su voluntad para implementar

“Hemos visto cómo ahora otras plataformas de redes sociales que se plantean los mismos temas que nosotros analizan nuestros procesos de toma de decisiones”.

Kristina Arriaga
FIDUCIARIA DEL CONSEJO
ASESOR DE CONTENIDO



las recomendaciones, como ha hecho Meta con muchas de nuestras propuestas. La asociación entre una empresa y un consejo asesor de contenido requiere la voluntad para aprender de ambas partes.

Como Consejo, esperamos que el enfoque descrito anteriormente proporcione un marco creíble para otras empresas tecnológicas que busquen beneficiarse de una supervisión independiente. A través de nuestro trabajo con Meta, ya hemos superado muchos de los obstáculos operativos y aprendido muchas de las lecciones asociadas a la creación de un organismo como este. Aunque cualquier enfoque tendría que adaptarse a las especificidades del trabajo de una empresa, nuestra experiencia podría ayudar a las empresas a beneficiarse de una supervisión independiente de forma más rápida y con menos costos.

Mientras pensamos sobre la mejor manera de responsabilizar a la industria tecnológica, animamos a todas las empresas a considerar la incorporación de una supervisión independiente en sus plataformas y servicios.

Ayudar a las empresas a adaptarse a la normativa emergente

Nos interesa trabajar con empresas que compartan nuestra creencia de que una gobernanza de contenidos transparente y responsable, supervisada por organismos independientes, es una parte esencial para la creación de un entorno en línea que respete la libertad de expresión y otros derechos humanos. Las empresas que tengan la voluntad de comprometerse de forma significativa con estas normas y estructuras para beneficiar a sus usuarios y a la sociedad en general, como ha hecho Meta, se ganarán la confianza de los usuarios y demostrarán a los reguladores que su intención es verdadera.

El año que viene, la nueva normativa tendrá nuevos requisitos para las empresas tecnológicas, lo que llevará a muchas a cambiar su enfoque. Los debates anteriores acerca de la regulación de las redes sociales han presentado la regulación gubernamental y la autorregulación del sector como una elección “exclusiva”, pero ahora estamos observando movimiento hacia la “corregulación”, en la que los elementos de un mecanismo regulador independiente están respaldados por la legislación. Aunque muchos actores desempeñarán su función en este panorama normativo emergente, creemos que nuestro enfoque independiente y foco en la transparencia y el trato justo hacia los usuarios pueden ser parte de la solución.

“No buscamos ser el consejo de todo el sector. En cambio, queremos compartir lo que hemos aprendido y trabajar con las empresas interesadas en constituir distintos organismos para establecer normas y supervisar la gobernanza de contenidos”.

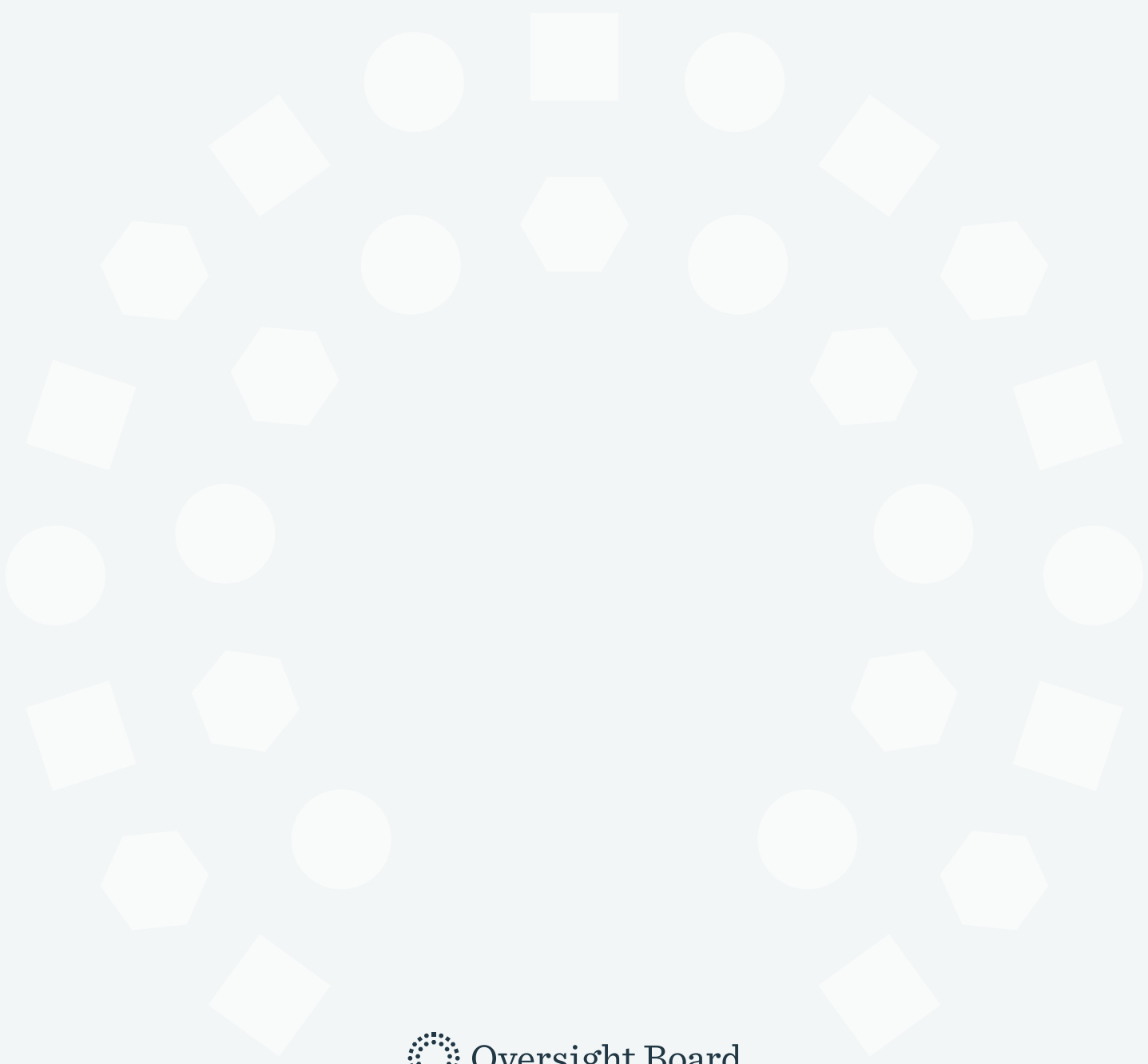
Thomas Hughes
DIRECTOR DEL CONSEJO
ASESOR DE CONTENIDO



Conclusión

En los mejores escenarios, las redes sociales pueden ser un catalizador sin igual para conseguir una conexión y conversación globales. Aprovechar estos beneficios, y al mismo tiempo contener sus daños, es una tarea abrumadora. Como entidad pionera, el Consejo asesor de contenido sigue aplicando las lecciones aprendidas para mejorarse a sí mismo. Creemos que las empresas de redes sociales toman decisiones de moderación de contenido de una manera más justa y fundamentada si basan esas decisiones en las normas internacionales sobre derechos humanos. Estamos listos para compartir lo que hemos aprendido hasta ahora con otras empresas y organizaciones que compartan nuestros objetivos de incrementar la transparencia y mejorar el trato que reciben las personas en Internet.





Oversight Board

www.oversightboard.com

© 2023 Oversight Board LLC