



Oversight Board

# 2021

INFORME  
ANUAL DE 2021  
DEL CONSEJO  
ASESOR DE  
CONTENIDOS



# Índice general

Prefacio	2
Prefacio de los copresidentes	4
Resumen ejecutivo	6
Aplicación de normas internacionales sobre derechos humanos a la moderación de contenidos a escala global: Artículo 19	9
Cronología de eventos clave	12

## Selección de casos

Información general	14
Casos presentados al Consejo	17
Casos que consideró el Comité de selección de casos	21

## Decisiones de casos

Información general	22
Tabla: decisiones publicadas en 2021	25
Enfoque de la decisión: primeras decisiones	27
Enfoque de la decisión: decisión relacionada con la suspensión del expresidente Trump	32
Enfoque de la decisión: garantizar el respeto del derecho de libertad de expresión	36
Enfoque de la decisión: protección de los usuarios frente a contenido perjudicial	41
Normas internacionales sobre derechos humanos en la toma de decisiones del Consejo	44
Preguntas formuladas por el Consejo a Meta como parte de las decisiones	50

## Comentarios del público

## Recomendaciones

Información general	56
Respuesta de Meta a nuestras recomendaciones	59
De compromisos a la acción: resultados para los usuarios	62
Petición de responsabilidades a Meta sobre la verificación cruzada	65
Participación y alcance	67

## Próximos pasos

2022 en adelante	69
------------------	----

## Anexo

Cómo respondió Meta a nuestras recomendaciones y cómo las implementó	71
--	----



# Prefacio

El Consejo asesor de contenidos tiene por objetivo mejorar el modo en que Meta trata a las personas y las comunidades de todo el mundo con la aplicación de las normas comunitarias de Facebook e Instagram para proteger la libertad de expresión y los derechos humanos. Para ello, cuenta con un proceso independiente para verificar la moderación de contenidos que realiza Meta y, de este modo, tomar decisiones vinculantes acerca de los principales desafíos en cuestiones de contenidos. De este modo, damos recomendaciones de políticas que obligan a Meta a optimizar sus reglas y a actuar de forma ética, transparente y equitativa con todos los usuarios.

En el año 2018 el número de usuarios de Facebook e Instagram creció hasta los dos mil millones de personas. La conexión entre personas ha reportado enormes beneficios, como por ejemplo, el acercamiento de seres queridos, el fortalecimiento de las comunidades ya existentes o la oportunidad de crear otras nuevas. Sin embargo, como contrapartida de esos beneficios, también han aparecido numerosos e importantes desafíos para la sociedad. Determinados contenidos suponen una amenaza para la libertad y la seguridad de las personas. Por su parte, algunos usuarios fomentan la desinformación, lo que supone una amenaza para la sociedad y la democracia. El lenguaje que incita al odio y los contenidos que promocionan lo que Meta identifica como “personas y organizaciones peligrosas” pueden dividir a las comunidades y socavar el valor que tienen las redes sociales como herramienta para mantener a las personas unidas.

Con el fin de iniciar conversaciones acerca de los pasos que debían darse para crear un sistema de gobierno que implementara un proceso de moderación de contenidos a escala mundial, Meta presentó en noviembre de 2018 una propuesta que permitía a las personas recurrir las decisiones relacionadas con contenidos a un órgano independiente y cuyas decisiones fuesen vinculantes. La creación de una entidad independiente de la compañía desde el punto de vista legal y financiero fue la consecuencia lógica del convencimiento de que Meta no debía tomar por su cuenta tantas decisiones importantes relacionadas con la libertad de expresión y la seguridad.

El Consejo asesor de contenidos se fundó atendiendo a la idea de promocionar los derechos e intereses de los usuarios a través de la transparencia y la aportación de una mayor responsabilidad, homogeneidad e imparcialidad en las decisiones relacionadas con contenidos.

En septiembre de 2019 se publicó un acta constitutiva que establecía la independencia institucional, los principios fundadores y los objetivos del Consejo. El acta constitutiva otorgaba al Consejo competencias para tomar decisiones vinculantes sobre la autorización para publicar determinados contenidos en Facebook e Instagram, decisiones que Meta tenía que implementar en un plazo de siete días tras su publicación. El acta también permitía al Consejo

emitir recomendaciones no vinculantes con el fin de mejorar el modo en que Meta trataba a las personas y las comunidades de todo el mundo. Junto con el acta constitutiva, se publicaron unos estatutos que establecían los procedimientos operativos del Consejo.

El Consejo asesor de contenidos consta de tres órganos que están entrelazados: Miembros del consejo, Fondo y Administración. Los fideicomisarios del Fondo son los responsables de salvaguardar la independencia del Consejo y de garantizar que este trabajo de manera eficaz. Los miembros del Consejo, entre los que se incluyen los cuatro copresidentes, se ocupan de seleccionar los casos y tomar decisiones al respecto, además de emitir recomendaciones para Meta y dirigir el trabajo de implementación. La Administración se compone de un equipo de personas que trabajan a tiempo completo para ayudar a los miembros del Consejo en su trabajo.

Los miembros del Consejo son líderes que tienen experiencia en el trabajo con situaciones complejas, como por ejemplo las relacionadas con los derechos humanos. Entre los miembros del Consejo se encuentran académicos, líderes de la sociedad civil, antiguos jueces, mandatarios de las Naciones Unidas y de entidades regionales especializadas en derechos humanos. Los miembros del Consejo han vivido en 27 países distintos, hablan unos 29 idiomas diferentes y lo integran el mismo número de hombres que de mujeres. El Fondo también está compuesto por el mismo número de hombres que de mujeres. La plantilla de Administración, pese a ser más reducida, también habla 28 idiomas, lo que viene a reforzar aún más el enfoque global del Consejo. Aproximadamente el 60% de la plantilla son mujeres y el 40% son hombres.

Para cumplir nuestros compromisos de responsabilidad y transparencia, hemos publicado diversos informes trimestrales sobre transparencia. La publicación del presente informe anual tiene por objetivo ofrecer al público una síntesis detallada de los casos que se han presentado ante el Consejo, un resumen de las decisiones y recomendaciones del Consejo e información general sobre comentarios del público. Por último, cabe decir que, por primera vez, este informe incluye un análisis que explica cómo las decisiones del Consejo han tenido en cuenta las normas internacionales sobre derechos humanos y las han aplicado en diversos casos representativos.

# Prefacio de los copresidentes



Catalina Botero-Marino, Jamal Greene, Michael McConnell, Helle Thorning-Schmidt  
*Copresidentes del Consejo asesor de contenido*

La idea a partir de la cual nació el Consejo (esta se basa en el hecho de que las compañías de redes sociales no tienen que tomar por sí mismas decisiones relacionadas con la moderación de contenidos) era sencilla de expresar, pero difícil de poner en práctica. Durante nuestro primer año, comenzamos a plasmar esta idea en la vida real.

De este modo, creamos sistemas de apelaciones independientes que se pusieron a disposición de miles de millones de usuarios de todo el mundo. Asimismo, implantamos un procedimiento de comentarios del público para que las personas pudieran tener voz en nuestras decisiones. Así fue como aprendimos, a pesar de las diferencias en nacionalidades, experiencias y puntos de vista, a deliberar sobre casos que no tenían respuesta fácil. En cada uno de los casos, se examinaba el contenido objeto del conflicto con respecto a las normas comunitarias de Facebook o Instagram relevantes y sin olvidar los valores propios de Meta. De este modo, evaluábamos si los principios se habían aplicado de manera correcta y homogénea, si las notificaciones y los procesos que habían recibido los usuarios eran adecuados y si las normas y directrices eran compatibles con las normas

internacionales sobre derechos humanos con los que Meta se había comprometido como entidad.

La organización del Consejo gira en torno al principio de que la libertad de expresión es un componente esencial de toda sociedad democrática, que debe respetarse para, de este modo, proteger los derechos humanos. Nuestro compromiso por la libertad de expresión está alineado con el pacto internacional relativo a los derechos civiles y políticos, que afirma que “toda persona tiene derecho a la libertad de expresión”. Por tanto, en todas nuestras decisiones se aprecia que las normas internacionales sobre derechos humanos son una importante fuente de autoridad para el Consejo.

En este primer informe anual, que abarca desde octubre de 2020 (fecha en la que comenzamos a aceptar apelaciones) hasta diciembre de 2021, se describen los progresos que se han realizado para optimizar el trato que dispensa Meta a usuarios y poblaciones afectadas de todo el mundo. La emergencia a causa de la pandemia global hizo aún más compleja la creación de una institución que fuese capaz de abordar esas tareas.

Entre los usuarios de Facebook e Instagram surgió una necesidad enorme de crear algún medio que les permitiera recurrir las decisiones de moderación de contenidos de Meta ante otra organización que fuese independiente de Meta. Prueba de ello es que durante este período, los usuarios presentaron **más de un millón de apelaciones** ante el Consejo para pedir, en su mayoría, la restauración de contenidos en publicaciones de Facebook o Instagram que supuestamente infringían las normas de Meta en materia de *bullying*, lenguaje de incitación al odio o de violencia e incitación. En abril de 2021 también se ofreció a los usuarios la posibilidad de apelar al Consejo asesor de contenidos para la eliminación de contenidos de Facebook o Instagram.

De este modo, en 2021, **emitimos decisiones con explicaciones detalladas públicas para 20 casos importantes** relacionados con cuestiones de lenguaje que incita al odio y la desinformación sobre COVID-19. Con esto, adoptamos un enfoque centrado en los derechos humanos para analizar las decisiones de moderación de contenidos y **recibimos unos 10.000 comentarios del público** que ayudaron a tomar las decisiones iniciales. En muchos más casos, el trabajo del Consejo hizo que la compañía optara por revertir de manera voluntaria decisiones de moderación de contenidos erróneas.

También publicamos un total de **86 recomendaciones para Meta**, que presionó a la compañía para aumentar la transparencia con respecto a sus políticas. Las respuestas de Meta a nuestras decisiones sobre casos y recomendaciones de políticas están comenzando a mejorar el tratamiento que la compañía dispensa a los usuarios. Ahora Meta proporciona a las personas que utilizan Facebook en inglés y que infringen las normas sobre lenguaje que incita al odio más detalles sobre qué es lo que han hecho mal. En determinadas ubicaciones, la compañía está implementando sistemas de mensajería para notificar a las personas la eliminación de contenidos como resultado de procesos de revisión humana o automatizada. Asimismo, la compañía se ha

comprometido a proporcionar nueva información sobre solicitudes de poderes públicos y datos acerca de su reconocimiento de interés periodístico en los informes de transparencia. Meta también se ha comprometido a traducir sus Normas Comunitarias a varios idiomas de la India, lo que significa que, una vez realizada la traducción, más de 400 millones de personas podrán leer las normas de Facebook en su lengua materna.

Pese al compromiso de Meta para implementar la mayor parte de las recomendaciones que se hicieron en 2021, nuestra siguiente tarea es garantizar que la compañía cumpla sus promesas para mejorar la experiencia de las personas que utilizan Facebook e Instagram. Por lo tanto, en este informe se aplica un nuevo enfoque basado en datos para realizar un seguimiento de la implementación de cada una de las recomendaciones por parte de la compañía. También estamos tratando de obtener de Meta nuevos datos que ayuden a comprender el impacto que tienen nuestras propuestas en los usuarios.

Tras un año de toma de decisiones, el Consejo ha comenzado a sentar las bases de un proyecto que generará cambios en Meta y que ayudará a la compañía a prestar un mejor servicio a las personas y comunidades afectadas por el extraordinario alcance que tienen sus plataformas de redes sociales. En 2022, seguiremos trabajando sobre esta base inicial. Asimismo, incorporaremos nuevos miembros al Consejo, abordaremos nuevas áreas de trabajo y seguiremos encaminando a Meta en la dirección correcta. Creemos que el éxito que ya ha obtenido el Consejo asesor de contenidos a la hora de exigir responsabilidades a Meta es una prueba fehaciente de la viabilidad de este, ya que proporciona un marco de trabajo autoregulatorio que ayudará a ampliar y mejorar nuestras operaciones de cara al futuro. De este modo, el Consejo formará parte del esfuerzo colectivo que realicen las compañías, las entidades gubernamentales, las instituciones académicas y la sociedad civil para modelar un futuro digital mejor y más seguro que beneficiará a las personas de todo el mundo.

# Resumen ejecutivo

**1,1m+**

presentados ante el Consejo por parte de los usuarios y de Meta

**Más de la mitad**

de nuestras decisiones guardan relación con países del sur global.



Análisis de las decisiones de moderación de contenidos de Meta según las

**normas internacionales sobre derechos humanos.**



Más de **8 de cada 10**

usuarios apela a la restauración de contenidos en relación con las reglas de Meta sobre bullying, lenguaje que incita al odio o violencia e incitación.

Publicación de un total de **20** decisiones

relacionadas con cuestiones de desinformación sobre el COVID y lenguaje que incita al odio.



META SE RETRACTÓ EN EL **70%**

de las decisiones de los casos: Meta se retractó de sus decisiones de moderación de contenidos en 14 ocasiones y se ratificó en 6 ocasiones.



Se han tenido en cuenta aproximadamente

**10.000**

comentarios del público de personas y organizaciones de todo el mundo para la toma de decisiones.



**86** recomendaciones emitidas para Meta.

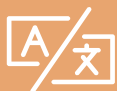
A continuación se detallan algunas de las recomendaciones del Consejo con las que Meta se ha comprometido:



**Proporcionar más detalles** a los usuarios a la hora de eliminar publicaciones con lenguaje que incita al odio.



Proporcionar más **informes acerca de las solicitudes de poderes públicos.**



Traducir normas a **los idiomas de más de 400 millones de personas.**



Adoptar un nuevo **protocolo de políticas de crisis** para gestionar la respuesta ante situaciones de crisis.



Implementar nuevos sistemas de mensajería en determinadas ubicaciones para notificar a los usuarios la eliminación de contenidos como resultado de procesos de **revisión humana o automatizada.**



En este informe se usa un **nuevo enfoque basado en datos** para realizar un seguimiento de la implementación y garantizar que Meta honre los compromisos adquiridos al aceptar las recomendaciones.



En **2/3** de las 86 recomendaciones, Meta acreditó la realización de progresos o la implementación de las recomendaciones, entrando en estas categorías las recomendaciones de **transparencia.**

**Objetivos para 2022 y para años venideros:**

Mantener conversaciones con Meta para **expandir nuestro ámbito de actuación**, lo que incluye la revisión de apelaciones de usuarios ante decisiones que afectan a grupos y cuentas.



**Incrementar el alcance de las partes interesadas** en Asia, América Latina, Oriente Medio y África.

## Control del cumplimiento de las responsabilidades de Meta por parte del Consejo

El Consejo está satisfecho con las tendencias observadas durante el primer año tras la colaboración con Meta; sin embargo, la compañía debe mejorar urgentemente su transparencia. Nos complace informar de que la compañía ha cumplido íntegramente sus compromisos en materia de decisiones de casos. Asimismo, ha aceptado la implementación de más de la mitad de las recomendaciones del Consejo y responde cada vez más a las consultas de este.

Sin embargo, existen cuestiones que siguen preocupando al Consejo en cuanto a la transparencia y la provisión de información relacionada con determinados casos y recomendaciones de políticas de Meta.

### Desde octubre de 2020 a diciembre de 2021, Meta...

- ✔ Implementó el 100% de las decisiones de casos y se comprometió a aplicar la mayor parte de nuestras recomendaciones.
- ✔ Respondió a una gran parte de las preguntas que se formularon como parte de las decisiones de casos a finales de 2021 hasta alcanzar una tasa del 94% en el cuarto trimestre de 2021.
- ✔ Mejoró la respuesta a nuestras recomendaciones, con una proporción de respuestas clasificadas como “completas” o “bastante completas” cada vez mayor en cada trimestre de 2021.
- ✘ No mostró disposición a la hora de proporcionar al Consejo información sobre su sistema de verificación cruzada.
- ✘ Perdió y no aplicó una importante excepción a una política por un período de tres años, tal como describe nuestra decisión acerca del aislamiento de Öcalan.
- ✘ Admitió que las decisiones originales tomadas en 51 de los 130 casos que seleccionó el Consejo eran incorrectas.



# Aplicación de normas internacionales sobre derechos humanos a la moderación de contenidos a escala global: Artículo 19

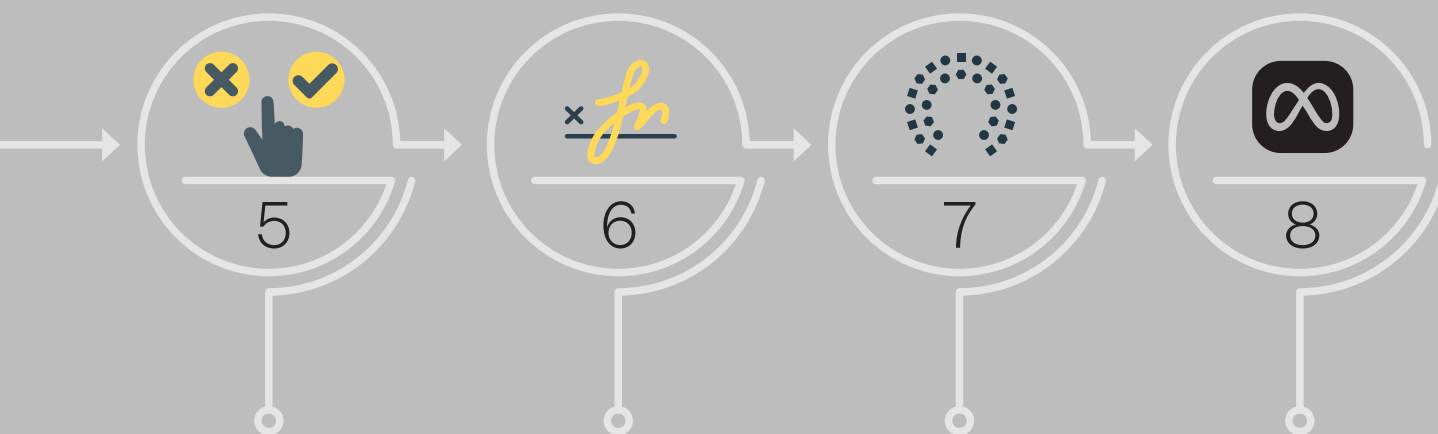
Uno de los temas más importantes que define el trabajo del Consejo de 2021 es la convicción de que Meta tomará decisiones de moderación de contenidos de manera más justa y ética si estas se basan en las normas internacionales sobre derechos humanos con las que se ha comprometido. Para ello, el acta constitutiva del Consejo asesor de contenidos establece que “se prestará especial atención al impacto de la eliminación de contenidos en cuanto a la posible vulneración de las normas sobre derechos humanos que protegen la libertad de expresión”. Esas normas incluyen el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos, según el cual “toda persona tiene derecho a la libertad de expresión [...]. El ejercicio de este derecho [...] puede estar sujeto a ciertas restricciones, que deberán, sin embargo, estar expresamente fijadas por la ley”. Este artículo incluye una prueba en tres partes que permite evaluar las restricciones de expresión:

- 1 ¿Cumple la restricción con el principio de **legalidad**? Aquí se evalúa si las normas en las que se basó Meta para tomar su decisión son accesibles y están lo suficientemente claras como para que los usuarios puedan comprenderlas y cumplirlas. Es importante que las normas sean claras para que las personas responsables de velar por su cumplimiento puedan tomar decisiones imparciales y coherentes.
- 2 ¿Tiene la restricción propuesta un **fin legítimo**? El Consejo ha analizado los objetivos de las restricciones que se aplican al derecho de expresión que reconoce el artículo 19 del ICCPR para evaluar si la norma en la que se basó la decisión persigue un objetivo compatible con los derechos.
- 3 ¿Puede considerarse que la restricción **es necesaria y proporcionada**? ¿Puede afirmarse que la eliminación del contenido fue el medio menos intrusivo para lograr el objetivo?

# Consideración de las apelaciones de usuarios por parte del Consejo<sup>1</sup>



<sup>1</sup> Gráfico: Consejo asesor de contenido. Fuente: herramienta de administración de casos del Consejo asesor de contenido.



### Decisión

El panel llega a una decisión sobre si permitir el contenido, si se sostiene o si se revoca a Meta.

### Aprobación

Se difunde entre todos los miembros del Consejo un borrador de la decisión para que se revise. La mayoría debe estar de acuerdo para que se haga pública una decisión.

### Publicación

Nuestra decisión se publica en la página-web del Consejo asesor de contenidos.

### Respuesta

En 2021, Meta disponía de 30 días desde la publicación de una decisión para implementar nuestra decisión y responder a nuestras recomendaciones. El último plazo se ha extendido a 60 días.

# Cronología de eventos clave



21 DE SEPTIEMBRE

**RESPUESTA A LAS REVELACIONES DE VERIFICACIÓN CRUZADA**

Siguiendo los informes de la prensa, empezamos a analizar si Meta estaba siendo completamente honesto con nosotros sobre el sistema de verificación cruzada.

15 DE JUNIO

**PRIMERA OPINIÓN CONSULTIVA SOBRE POLÍTICAS ACEPTADA**

Aceptación de la opinión consultiva sobre políticas al compartir información residencial privada. Las opiniones consultivas sobre políticas nos permiten sugerir a Meta cómo podría mejorar una política concreta.

21 DE OCTUBRE

**INFORMES DEL PRIMER TRIMESTRE**

Publicación de los informes sobre transparencia del primer trimestre teniendo en cuenta que el equipo relevante en Meta no fue completamente honesto con la verificación cruzada.

8 DE JULIO

**DECISIÓN SOBRE EL AISLAMIENTO DE ÖCALAN**

Publicación de la decisión acerca del aislamiento de Öcalan; cabe destacar que Meta perdió y no aplicó una importante excepción a una política por un período de tres años.

26 DE MAYO

**PROTESTAS A FAVOR DE NAVALNY**

Publicación de la primera decisión en la que revocamos la decisión de Meta atendiendo a sus responsabilidades sobre derechos humanos, aunque la eliminación era coherente con las reglas de Facebook.

# Selección de casos



Entre octubre de 2020 y diciembre de 2021 se presentaron en total

**1.152.181** casos ante el Consejo

CON

**47**

de Meta

Cada día, el Consejo recibe una media de

**2.649**

casos.



Más de

**8 de cada 10**

usuarios apela a la restauración de contenidos en relación con las reglas de Meta sobre bullying, lenguaje que incita al odio o violencia e incitación.



de los casos presentados procedían de EE.UU., Canadá y Europa.

EN  
**51**

DE LOS  
**130**

casos que seleccionó el Consejo, Meta admitió que la decisión que tomó originalmente en cuanto al contenido era incorrecta.



## Información general

Los usuarios de Facebook e Instagram pueden impugnar las decisiones de Meta si apelan contra el contenido ante el Consejo. Entre octubre de 2020, fecha en la que el Consejo comenzó a aceptar apelaciones de los usuarios, y diciembre de 2021 (fecha límite para el ámbito del presente informe), se recibieron más de 1,1 millones de solicitudes de usuarios de revisión independiente de decisiones de Meta sobre moderación de contenidos. Meta también remitió 47 casos al Consejo. Desde octubre de 2020 a diciembre de 2021, el Consejo recibió un promedio de 2.649 casos diarios.

El volumen de casos presentados es una clara muestra de la importancia del trabajo del Consejo para los usuarios. Tanto en 2020 como en 2021, hemos dado prioridad a los casos con capacidad para afectar a numerosos usuarios de todo el mundo, a los más importantes para el discurso público o a los que planteaban importantes preguntas acerca de las políticas de Meta. Con el fin de abordar los problemas exclusivos de personas de países específicos, también se seleccionaron casos de distintas regiones de todo el mundo. Además, también se seleccionaron varios casos estrechamente relacionados con la aplicación de las normas internacionales sobre derechos humanos a la moderación de contenidos a escala mundial.

El Consejo también **tuvo en cuenta a los usuarios** y dio prioridad a las apelaciones de casos relacionados

con problemas recurrentes. Eso implica que, a pesar de seleccionar solo un caso relacionado con un problema determinado para su revisión, el Consejo puede abordar problemas comunes que afecten a un grupo mayor de personas, incluidas aquellas cuyos casos no se hayan seleccionado. Por ejemplo, cuando vimos que Meta eliminó numerosas publicaciones que hacían referencia a varias figuras nazis a pesar de que el contenido no apoyaba ni promocionaba a ningún individuo u organización peligrosa, seleccionamos el caso de la *cita nazi* para examinar un importante problema que afectaba a un gran número de usuarios.

Con el fin de garantizar que contamos con una amplia gama de opiniones como base para el proceso de selección de casos, el Consejo nombra de manera periódica a un nuevo Comité de selección de casos para tratar diferentes problemas comunes a grandes grupos de usuarios. Hasta la fecha, en línea con los criterios generales de selección de casos, los comités han dado prioridad a aquellos en los que se ha moderado contenido mediante sistemas automáticos de Meta, sobre todo en países en los que los usuarios informaron de una falta de procedimiento de revisión humana en el idioma materno; casos y contenidos relacionados con el presunto apoyo o promoción de “personas y organizaciones peligrosas” y casos que planteaban importantes problemas de libertad de prensa, sobre todo de periodistas en peligro de “daño inminente” por su presencia en zonas de conflicto.



## LECCIONES APRENDIDAS

A continuación se describen los obstáculos y retos operativos que tuvo que superar el Consejo a la hora de seleccionar los casos:

### **Adaptación al derecho de libre expresión de los usuarios para eliminar el contenido propio tras la selección de casos**

Ejemplo: al poco tiempo de anunciar los primeros casos seleccionados en diciembre de 2020, un usuario decidió eliminar una parte de su contenido que estaba relacionada con unos tweets que había publicado el ex primer ministro de Malasia, el Dr. Mahathir Mohamad. Esta eliminación impidió que el Consejo aceptara el caso para su revisión. En una situación similar que se dio a finales de 2021, un usuario eliminó contenido que hacía referencia a un periodista que estaba trabajando en Afganistán durante la presencia de las tropas de Estados Unidos en el país y tras su retirada. En ambos casos, pese a que la eliminación de contenidos impidió al Consejo abordar áreas de interés común, se respetaron los derechos de los usuarios a retirar contenidos de las publicaciones fueran cuales fueran sus razones. Puesto que estas

situaciones pueden darse cada cierto tiempo, el Consejo ha decidido que anunciará y proporcionará explicaciones de inmediato siempre que se retire un caso debido a actuaciones de los usuarios.

### **Aportación de mayor transparencia ante casos de eliminación de contenido debido a la acción del opaco “sistema de faltas” de Meta**

En el segundo trimestre de 2021, se asignaron dos casos a un panel sin anunciarlo a los usuarios debido a las acciones de cumplimiento normativo que emprendió Meta. De este modo, dejó de anunciarse un caso sobre los orígenes del COVID-19 tras la eliminación por parte de Meta de la página que hospedada el contenido por recibir demasiadas “faltas”. En otro caso relacionado con el “misgendering”, Meta eliminó la página que hospedaba la publicación por haber recibido demasiadas “faltas”.

# Casos presentados al Consejo<sup>2</sup>

Desde que comenzamos a aceptar casos en octubre de 2020, el volumen de casos presentados ante el Consejo hasta el tercer trimestre de 2021 ha experimentado un continuo aumento hasta registrar una leve reducción en el cuarto trimestre de 2021. No obstante, esto indica una prevalencia global de problemas de moderación de contenido entre los usuarios de Facebook e Instagram.

El número de apelaciones registrado entre el segundo y el tercer trimestre de 2021 aumentó en un 64%. Es probable que este crecimiento se deba a las mejoras que realizó Meta en el proceso de apelación ante el Consejo desde la aplicación móvil de Facebook, así como a la mayor presencia del Consejo. Desde el 30 de agosto de 2021, los usuarios establecieron un récord de presentación de casos en una sola semana con la presentación de más de 32.000 apelaciones individuales al Consejo.

De los casos presentados, el 99% estaba relacionado con contenido de Facebook,<sup>3</sup> mientras que solo el 1%

## Casos estimados presentados al consejo por trimestre

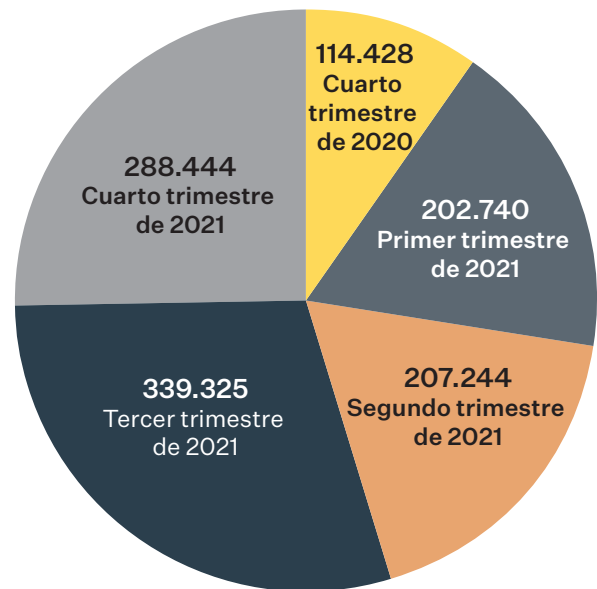


Gráfico: Consejo asesor de contenidos. Fuente: herramienta de administración de casos del Consejo asesor de contenidos.



- Debido a las limitaciones de la herramienta de administración de casos a través de la cual se presentan los casos al Consejo y los datos que tenemos disponibles actualmente, en algunas ocasiones el equipo de selección de casos contó manualmente los casos que se presentaron a medida que fueron apareciendo en el CMT. El desglose de esos números también puede estar basado en ejemplos. Como tales, esos números deben considerarse estimaciones.
- Esta figura solo cubre el período que va entre el segundo y el cuarto trimestre de 2021, ya que únicamente se pudieron obtener datos de este.

## Casos estimados presentados al consejo por norma comunitaria (oct 2020 - dic 2021)

(solo incluye las apelaciones de usuarios para restaurar contenido en Facebook e Instagram)

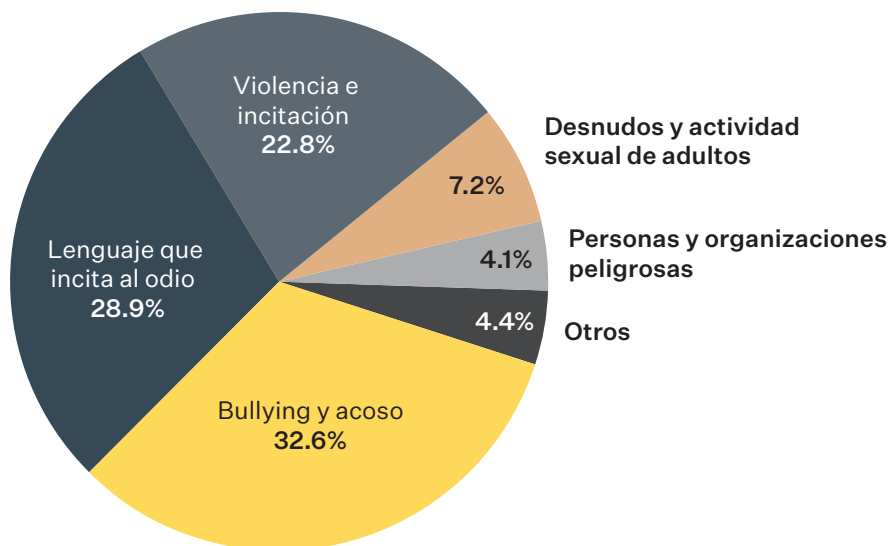


Gráfico: Consejo asesor de contenidos. Fuente: herramienta de administración de casos del Consejo asesor de contenido.

hacia referencia a contenido de Instagram. A partir del segundo trimestre de 2021, tras proporcionar a los usuarios medios para apelar al Consejo y eliminar contenidos *de otras personas*, el 94% de las apelaciones de usuarios se presentaron para *restaurar* contenidos en Facebook o Instagram, mientras que el 6% restante eran apelaciones para *eliminar* contenidos de las plataformas.

La inmensa mayoría de las apelaciones de los usuarios para restaurar contenido (84%) estuvo relacionada con tres normas comunitarias de Facebook: *bullying* y acoso, lenguaje que incita al odio y violencia e incitación. La mayor parte del resto de las apelaciones

de restauración de contenidos correspondió a casos relacionados con desnudos y actividad sexual de adultos, así como a personas y organizaciones peligrosas.

A modo de contexto adicional para estas cifras, cabe destacar que desde el cuarto trimestre de 2020 hasta el cuarto trimestre de 2021, las acciones de Meta en la mayor parte de los contenidos (eliminación de contenidos y otras acciones, como, por ejemplo, la visualización de pantallas de advertencia) se emprendieron en base a las normas comunitarias de desnudos y actividad sexual de adultos, seguidas de contenido gráfico y violento y lenguaje que incita

## Casos estimados presentados al consejo por norma comunitaria por trimestre (solo incluye las apelaciones de usuarios para restaurar contenido en Facebook e Instagram)

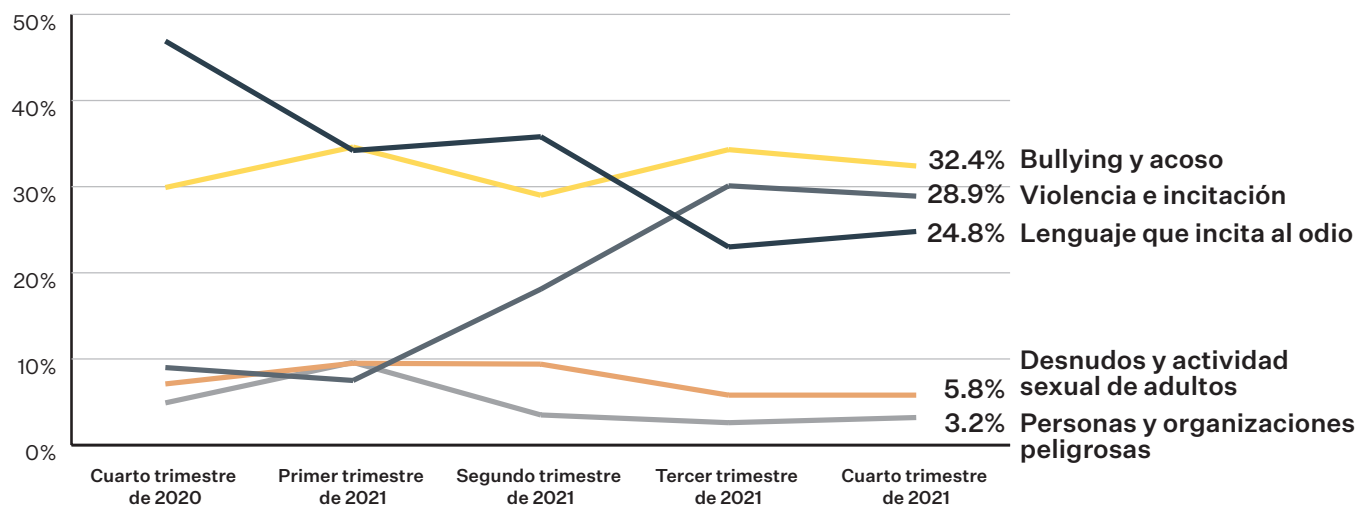


Gráfico: Consejo asesor de contenidos. Fuente: herramienta de administración de casos del Consejo asesor de contenido.

## Casos estimados presentados al consejo por regiones seleccionadas de los usuarios (oct 2020 - dic 2021)

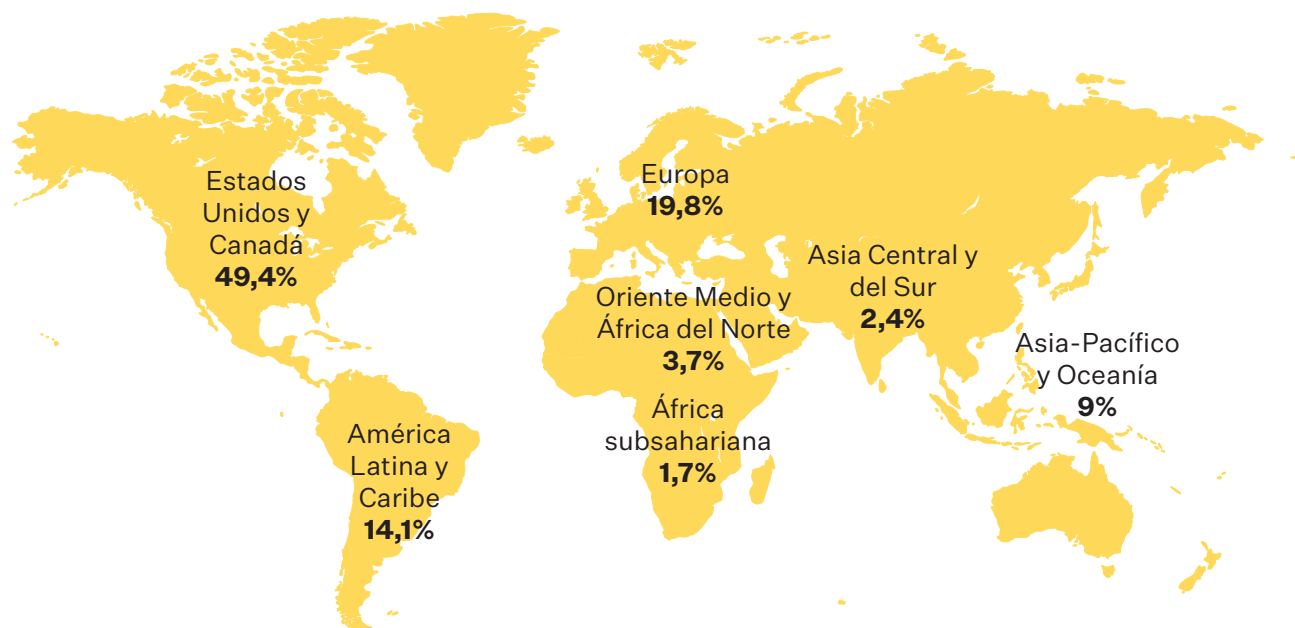


Gráfico: Consejo asesor de contenidos. Fuente: herramienta de administración de casos del Consejo asesor de contenido.

al odio. Sin embargo, las normas comunitarias para las que Meta recibió más apelaciones durante este periodo fueron las de lenguaje que incita al odio seguidas de *bullying* y acoso, y desnudos y actividad sexual de adultos.<sup>4</sup> Por el contrario, durante este mismo periodo, las dos normas comunitarias para las que el Consejo recibió la mayoría de las apelaciones de restauración de contenidos por parte de los usuarios fueron *bullying* y acoso, y lenguaje que incita al odio.

El análisis de los datos de cada trimestre sobre apelaciones de usuarios para la restauración de contenido en las dos plataformas revela importantes tendencias. Aunque la norma comunitaria de *bullying* y acoso representó aproximadamente un tercio de las apelaciones, otras normas comunitarias experimentaron importantes cambios durante el año. La tasa de casos relacionados con las reglas de Facebook sobre violencia e incitación se ha triplicado de un año para otro al pasar del 9% en el cuarto trimestre de 2020 al 29% en el cuarto trimestre de 2021. Durante ese mismo periodo, la tasa de casos relacionados con el lenguaje que incita al odio se redujo prácticamente a la mitad al pasar de un 47% (cuarto trimestre de 2020) a solo un 25% (cuarto trimestre de 2021).

Más de dos tercios de las apelaciones de los usuarios provinieron del Norte Global, con un 49% del total de las apelaciones procedentes de Estados Unidos y Canadá, y un 20% de Europa. El 14% de las apelaciones procedieron de América Latina y el Caribe, el 9% de Asia-Pacífico y Oceanía, el 4% de Oriente Medio y África del Norte, el 2% de Asia Central y del Sur y el 2% de África subsahariana.

Sabemos que esta distribución no refleja la extensión de los usuarios de Facebook e Instagram en el mundo. En el año 2019, por ejemplo, solo seis de los 20 países con más usuarios de Facebook estaban en Europa y América del Norte, mientras que la India es el país con más usuarios de Facebook e Instagram. Las bajas cifras de apelaciones de usuarios de fuera de Europa y Estados Unidos y Canadá también podrían indicar que muchos de los usuarios que usan Facebook e Instagram en el resto del mundo no conocen la posibilidad de apelar ante el Consejo las decisiones de moderación de contenidos de Meta. Tampoco creemos que la distribución de datos de apelaciones de este mapa refleje la distribución real de problemas de moderación de contenidos en todo el mundo. En todo caso, tenemos razones para pensar que los usuarios de Asia, África y Oriente Medio experimentan más

4 Fuente: Informes del Sistema de cumplimiento de normas y de Normas comunitarias de Meta (CSER). Las Normas comunitarias mencionadas aquí excluyen las áreas de las políticas fuera del alcance del Consejo y las que no se mencionen en el CSER.

problemas (y no menos) con las plataformas de Meta que en otros lugares del mundo. Nuestras decisiones, que también abarcaron publicaciones de la India y Etiopía, plantean dudas sobre si Meta realmente ha invertido suficientes recursos en la moderación de contenidos en otros idiomas distintos del inglés.

Además de expandir nuestro alcance en regiones fuera de Europa, Estados Unidos y Canadá, también se están considerando y seleccionando numerosos casos de fuera del Norte Global para abordar las preocupaciones de los usuarios de todo el mundo. Más de la mitad de los 278 casos que consideró el Comité de selección de casos procedían de fuera de Europa, Estados Unidos y Canadá, y más de la mitad de las 20 decisiones publicadas en 2021 estaban relacionadas con países del sur global.

## Casos remitidos por Meta

Además de las apelaciones de usuarios de Facebook e Instagram, Meta también puede remitir al Consejo casos importantes y complejos.

De los 47 casos que Meta remitió al Consejo hasta diciembre de 2021, 41 estaban relacionados con contenidos compartidos en Facebook y seis con contenidos de Instagram. 29 de los 47 casos que remitió Meta estaban relacionados con contenidos que estaban disponibles en Facebook e Instagram y que aún no se habían considerado como infracciones de las normas de la compañía.

Los 18 casos restantes correspondieron a contenidos que Meta eliminó de Facebook o Instagram por presunta infracción de las siguientes normas comunitarias: violencia e incitación (seis casos), lenguaje que incita al odio (tres casos), personas y organizaciones peligrosas (tres casos), explotación sexual infantil, abusos y desnudos (tres casos), bullying y acoso (dos casos) y desnudos y actividad sexual de adultos (un caso).

Aproximadamente dos tercios de los casos que remitió Meta al Consejo estaban relacionados con contenido procedente de Europa, Estados Unidos y Canadá, con 30 de los 47 casos. Meta remitió 17 casos correspondientes a lugares del resto del mundo: cuatro de América Latina y el Caribe, cuatro de Oriente Medio y África del Norte, tres de Asia Central y del Sur, tres de Asia-Pacífico y Oceanía y tres de África subsahariana.

## Casos remitidos a Meta por región



# Casos que consideró el Comité de selección de casos

Para abordar los problemas de los usuarios de Facebook e Instagram de todo el mundo, el Consejo selecciona una amplia variedad de casos en muchos idiomas distintos. De este modo, los 278 casos que consideró el Comité de selección de casos hasta diciembre de 2021 abarcan más de 70 países: desde Fiyi a Chad, pasando por Trinidad y Tobago. Estos casos incluyen contenidos en 38 idiomas distintos.

A veces, tras preseleccionar los casos que se someterán a la revisión del Consejo por parte del Comité de selección de casos, Meta determina que la decisión inicial que tomó con respecto a determinados ciertos era incorrecta. Esto es lo que se conoce como un “error del sistema de cumplimiento de normas”.

El gráfico de esta página muestra los casos que Meta identificó como errores del sistema de cumplimiento de normas por cada trimestre. Tanto en el cuarto trimestre de 2020 como en el segundo trimestre de 2021, Meta estimó que la decisión que tomó para un

tercio de los casos seleccionados era incorrecta. En el primer trimestre de 2021, este índice era del 44%, mientras que en el tercer trimestre de 2021, del 46%. El Consejo no preseleccionó ningún caso durante el cuarto trimestre de 2021.

De los 130 casos que destacó el Consejo hasta diciembre de 2021, Meta identificó que la decisión original que había tomado inicialmente con respecto al contenido era incorrecta en 51 ocasiones. A pesar de que se trata solo de una pequeña muestra y de que el Consejo busca de manera intencionada los casos más complejos, resulta preocupante constatar que Meta solo considera que su decisión fue incorrecta en menos de 4 de cada 10 casos. Este elevado índice de errores plantea preguntas más amplias acerca de la precisión de la moderación de contenidos de Meta y del proceso de apelación que aplica Meta antes de que los casos lleguen al Consejo.

## Errores del sistema de cumplimiento de normas según el número de casos preseleccionados

- Identificado como error del sistema de cumplimiento de normas
- No identificado como error del sistema de cumplimiento de normas

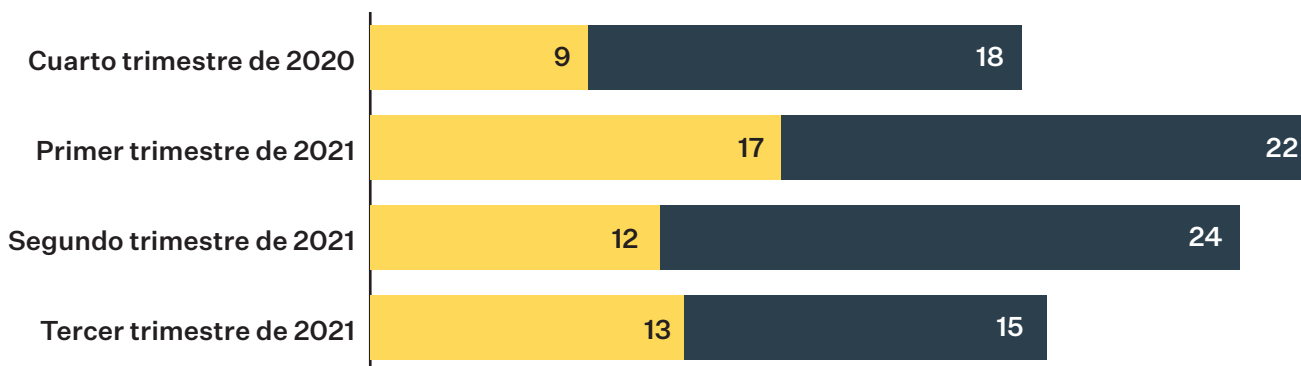


Gráfico: Consejo asesor de contenidos. Fuente: herramienta de administración de casos del Consejo asesor de contenido.

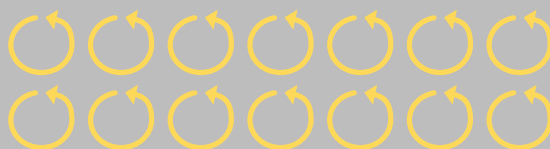
# Decisiones de casos



En 2021 se publicaron unas

**20**

decisiones



Meta se retractó en **14** ocasiones y se ratificó en **6** ocasiones.



Las decisiones contemplaron

**16** casos de usuarios y **4** de Meta.



**17** casos estaban relacionados con Facebook y **3** casos con Instagram.

El Consejo necesitó un plazo medio de **83** días para tomar una decisión con respecto a cada caso e implementarla.



**Más de la mitad**

de nuestras decisiones guardan relación con países del sur global.



El Consejo formuló **313** preguntas a Meta como parte de su proceso de revisión de casos.



Análisis de las decisiones de moderación de contenidos de Meta según las **normas internacionales sobre derechos humanos.**





# Información general

El cometido del Consejo asesor de contenidos es el de tomar decisiones independientes y vinculantes con respecto a los problemas de contenido más complejos que afectan a los usuarios de Facebook e Instagram. Las primeras 20 decisiones completas, así como las recomendaciones de políticas relacionadas, revelaron importantes problemas acerca del trato que proporciona la compañía a los usuarios de todo el mundo. El Consejo no solo revierte las decisiones de Meta cuando así lo estima necesario, sino que también recomienda cambios de políticas que mejoran el tratamiento que reciben los usuarios y que proporcionan mayor transparencia. En conjunto, en los casos en los que las políticas de la compañía son confusas o incoherentes, o cuando no se aplican, estas acciones se inclinan a favor de los usuarios. A continuación se detallan los temas que abordaron las decisiones que tomó el Consejo:

- **Protección de la libertad de expresión y de otros derechos humanos.** Como compañía que proporciona la principal red de comunicación de miles de millones de personas, Meta considera que la voz de sus usuarios es de “vital importancia”. Sin embargo, la libertad de expresión es un derecho que debe limitarse cuando sea necesario para proteger a los demás usuarios de perjuicios reales. Conciliar estas cuestiones es, sin duda, la responsabilidad más importante y compleja del Consejo.
- **Consideración del contexto.** A menudo, nuestras decisiones dependieron del contexto de cada situación. El contexto abarca desde el significado

de las palabras empleadas o de la publicación al completo hasta las circunstancias en las que se publicó el contenido, como por ejemplo la presencia de conflictos violentos o protestas políticas.

- **Claridad con los usuarios.** El Consejo tomó decisiones en relación con un caso en el que Meta no indicó a un usuario la norma comunitaria que había infringido, otro caso en el que la compañía no revisó la apelación de un usuario y otro caso en el que Meta incumplió una parte de las instrucciones de una política durante tres años. También se han tomado tres decisiones con respecto a contenidos de Instagram que pusieron de relieve que la relación entre las reglas de Facebook e Instagram no está clara para los usuarios.
- **Tratamiento imparcial a los usuarios con independencia del lugar en el que se encuentren.** Nuestras decisiones han sacado a relucir que las reglas de Meta de Facebook e Instagram no están disponibles en todos los idiomas de los usuarios. También se han planteado dudas sobre si Meta realmente invierte suficientes recursos en la moderación de contenidos en otros idiomas distintos del inglés.
- **Verificación de la correcta aplicación de las normas de Meta.** Nuestras decisiones indican claramente que Meta debe mejorar la precisión de sus procesos humanos y automatizados de moderación de contenidos y debe ser más transparente con el Consejo y el público acerca de las razones por las que se producen los errores.

## LECCIONES APRENDIDAS

### **Revisión de los plazos en consonancia con las realidades operativas**

Con los borradores de las primeras decisiones, nos percatamos rápidamente de que el plazo original de 90 días a contar desde la última decisión de Meta sobre un caso era insuficiente. Con el fin de poder contar con la misma cantidad de tiempo para la deliberación en todos los casos, en marzo optamos por cambiar el punto a partir del cual comenzaban a contar los 90 días de plazo. De este modo, el inicio del plazo se fijó en el momento en que se asignaba el caso a un panel. Cuando el consejo desarrolló sus procesos para la toma de decisiones y comenzó a plantear más preguntas a Meta, el inicio del plazo de 90 días volvió a modificarse en noviembre y se estableció la fecha de anuncio del caso como punto de partida.

### **Mejora de nuestras normas y procesos**

En las deliberaciones de los primeros casos, el Consejo tuvo que proporcionar respuestas para varias cuestiones complejas. Por ejemplo, a la hora de examinar un determinado contenido, ¿debe aplicarse toda la información recabada a través del proceso de revisión del caso o debe evaluarse si la decisión de Meta fue una respuesta razonable a la luz de las circunstancias y la información disponible en el momento? ¿Cómo debe el Consejo aplicar las normas internacionales sobre derechos humanos desarrolladas para los estados a las decisiones tomadas en plataformas de redes sociales? Esta cuestión se aborda en mayor profundidad en la sección sobre derechos humanos del presente capítulo.

# TABLA: DECISIONES PUBLICADAS EN 2021

	PLATAFORMA	FUENTE	NORMA COMUNITARIA	PAÍSES	DECISIÓN DEL CONSEJO
<b>Publicación de un usuario de Myanmar/Birmania acerca de los musulmanes</b> Núm. de caso: 2020-002-FB-UA		Usuario		Myanmar/ Birmania, Francia, China	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Armenios en Azerbaiyán</b> Núm. de caso: 2020-003-FB-UA		Usuario		Armenia, Azerbaiyán	<b>Ratificación</b> de la decisión de Meta de eliminar el contenido
<b>Síntomas del cáncer de mama y desnudos</b> Núm. de caso: 2020-004-IG-UA		Usuario		Brasil	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Cita nazi</b> Núm. de caso: 2020-005-FB-UA		Usuario		Estados Unidos	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Supuesta cura para el COVID</b> Núm. de caso: 2020-006-FB-FBR		Meta		Francia	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Protestas en la India contra Francia</b> Núm. de caso: 2020-007-FB-FBR		Meta		La India, Francia	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Suspensión del expresidente Trump</b> Núm. de caso: 2021-001-FB-FBR		Meta		Estados Unidos	<b>Ratificación</b> de la decisión de Meta de eliminar el contenido
<b>Representación de Zwarte Piet</b> Núm. de caso: 2021-002-FB-UA		Usuario		Países Bajos	<b>Ratificación</b> de la decisión de Meta de eliminar el contenido
<b>Preocupación punyabí sobre la RSS en la India</b> Núm. de caso: 2021-003-FB-UA		Usuario		La India	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Protestas a favor de Navalny en Rusia</b> Núm. de caso: 2021-004-FB-UA		Usuario		Rusia	<b>Revocación</b> de la decisión de Meta de eliminar el contenido



Personas y organizaciones peligrosas



Personas y organizaciones peligrosas



Lenguaje que incita al odio



Desnudos y actividad sexual de adultos



Bullying y acoso

	PLATAFORMA	FUENTE	NORMA COMUNITARIA	PAÍSES	DECISIÓN DEL CONSEJO
<b>Meme “dos botones”</b> Núm. de caso: 2021-005-FB-UA		Usuario		Armenia, Turquía, Estados Unidos	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Aislamiento de Öcalan</b> Núm. de caso: 2021-006-IG-UA		Usuario		Turquía, Estados Unidos	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Bot de Myanmar</b> Núm. de caso: 2021-007-FB-UA		Usuario		Myanmar/ Birmania	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Confinamientos por COVID en Brasil</b> Núm. de caso: 2021-008-FB-FBR		Meta	Ninguna	Brasil	<b>Ratificación</b> de la decisión de Meta de mantener el contenido
<b>Publicación compartida de Al Jazeera</b> Núm. de caso: 2021-009-FB-UA		Usuario		Israel, Egipto	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Protestas en Colombia</b> Núm. de caso: 2021-010-FB-UA		Usuario		Colombia	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Insultos en Sudáfrica</b> Núm. de caso: 2021-011-FB-UA		Usuario		Sudáfrica	<b>Ratificación</b> de la decisión de Meta de eliminar el contenido
<b>Cinturón Wampum</b> Núm. de caso: 2021-012-FB-UA		Usuario		Canadá	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Bebida de ayahuasca</b> Núm. de caso: 2021-013-IG-UA		Usuario		Brasil	<b>Revocación</b> de la decisión de Meta de eliminar el contenido
<b>Presunto delito en Raya Kobo</b> Núm. de caso: 2021-014-FB-UA		Usuario		Etiopía	<b>Ratificación</b> de la decisión de Meta de eliminar el contenido



Personas y organizaciones peligrosas



Lenguaje que incita al odio



Bienes regulados

## Primeras decisiones



Hasta el momento, [Mark] Zuckerberg era el responsable de tomar las decisiones con respecto al contenido... Sin duda este es un día histórico. Es la primera vez que la moderación de contenido se realiza fuera de Facebook y que Facebook se ve obligada a acatar nuestras decisiones.”

Helle Thorning-Schmidt  
*Copresidente del Consejo asesor de contenido*



El 28 de enero de 2021, el Consejo asesor de contenidos publicó cinco decisiones basadas en las deliberaciones que se produjeron desde la selección de los primeros casos de entre 20.000 apelaciones. Los miembros del Consejo deliberaron según el proceso que se describe en nuestros estatutos: cada caso se asignó a un panel de cinco miembros con al menos un miembro de la región implicada en el contenido y con una representación de género mixta. Cuando el panel tomó una decisión, se proporcionó a todos los miembros del Consejo un borrador para la recopilación de comentarios y la posterior votación por parte de todo el Consejo. Los casos abarcaron contenidos de cuatro continentes: Asia, Europa, América del Norte y América del Sur.

De los cinco primeros casos, las decisiones de Meta de eliminar publicaciones se revocaron en cuatro de los casos y se ratificó en uno. Los casos guardaban relación con una amplia variedad de problemas. En uno de ellos los miembros del Consejo examinaron si, en el contexto de un conflicto armado, Meta tenía derecho a eliminar una publicación admisible en cualquier otra circunstancia por el hecho de contener insultos que incitaban al odio. En otro, se examinó si una publicación acusada de propagar desinformación acerca del COVID-19 podía ocasionar perjuicios inminentes. Otra decisión sentó un importante precedente al poner de relieve la incapacidad de los sistemas de cumplimiento de normas automatizados de Meta para distinguir imágenes permitidas de pezones femeninos al descubierto en una campaña de concienciación contra el cáncer de mama.



## Síntomas del cáncer de mama y desnudos

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de enero de 2021	Brasil	Desnudos y actividad sexual de adultos

En octubre de 2020, un usuario de Brasil publicó una imagen en Instagram con un título en portugués cuyo objetivo era concienciar acerca del cáncer de mama. Un sistema automatizado encargado del cumplimiento de la norma comunitaria de Facebook en cuanto a desnudos y actividad sexual de adultos eliminó la publicación, ya que la compañía sostenía que la norma también era aplicable a Instagram. El Consejo revocó la decisión de Meta porque la norma comunitaria contempla, como excepción, la exhibición de desnudos en los casos en los que el usuario pretenda “concienciar acerca de una causa o con fines educativos o médicos” y permite, de manera específica, la publicación de imágenes de pezones femeninos al descubierto para “incrementar la concienciación acerca del cáncer de mama”.

Tras eliminar el contenido, el usuario apeló la decisión a Meta. En declaraciones públicas, Meta había comentado anteriormente que no siempre sería capaz de ofrecer a los usuarios la opción de apelar debido a la reducción temporal de su capacidad de revisión a causa del COVID-19. Además, Meta también declaró que no todas las apelaciones se someterían a procesos de revisión humana.

El Consejo rechazó el argumento de Meta que afirmaba que el caso era “irrelevante” debido a que, al restaurar la publicación, el usuario y Meta dejaban de estar en desacuerdo acerca de si el contenido debía seguir publicado. El Consejo también dejó claro que tiene autoridad para revisar casos de los usuarios incluso cuando Meta opte por rectificar sus errores a posteriori y restaure el contenido. Por último, el Consejo recomendó ofrecer a los usuarios la posibilidad de apelar las decisiones que tomen los sistemas automatizados para que se realice una revisión humana cuando se les informe de que su contenido ha infringido la norma comunitaria de Facebook sobre desnudos y actividad sexual de adultos.





## Cita nazi

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de enero de 2021	Estados Unidos	Personas y organizaciones peligrosas

En octubre de 2020, un usuario de Estados Unidos publicó una cita en inglés que se atribuía incorrectamente a Joseph Goebbels, el ministro de Propaganda del Reich en la Alemania nazi. La cita, en inglés, afirmaba que no tenía ningún sentido apelar a los intelectuales, ya que estos no pueden convertirse, y que, en cualquier caso, la ley del más fuerte es la que prevalece. La cita también afirmaba que los argumentos deben apelar a las emociones y los instintos. Meta eliminó la publicación por infringir la norma comunitaria de Facebook sobre personas y organizaciones peligrosas. Por su parte, el usuario mantuvo ante el Consejo que su intención era la de establecer una comparación entre el sentimiento plasmado en la cita y la presidencia de Estados Unidos: el presidente Donald Trump.

Al revisar el caso, el Consejo determinó que dado que la cita no apoyaba la ideología del partido nazi ni los actos de odio y violencia del régimen, no infringía la norma comunitaria de Facebook. En este caso, el Consejo se percató de que el planteamiento de Meta, según el cual los moderadores de contenidos debieron revisar el contenido sin tener en cuenta todo el contexto, tuvo como consecuencia una restricción innecesaria de la libertad de expresión del usuario. El Consejo también recomendó a Meta que proporcionara explicaciones y ejemplos de la aplicación de los términos clave que se usan en la política de personas y organizaciones peligrosas, incluidas las definiciones de los términos “elogio”, “apoyo” y “representación”. Estos deben coincidir con las definiciones que se usan en los estándares de implementación internos de Facebook.



## Supuesta cura para la COVID

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de enero de 2021	Francia	Violencia e incitación

Otra decisión se tomó en un caso en el que la desinformación de carácter sanitario llegó al punto de contribuir al riesgo de perjuicio físico inminente. En octubre de 2020 un usuario publicó una declaración en francés sobre la decisión de un órgano de la sanidad pública francesa (la Agence Nationale de Sécurité du Médicament o agencia francesa responsable de regular productos de salud) de no autorizar el uso combinado de hidroxycloquina y azitromicina como tratamiento contra el COVID-19. Ese usuario hacía referencia a esta combinación de fármacos como “medicamento inocuo” y la identificaba como “cura”. Meta eliminó la publicación por considerar que infringía su política sobre violencia e incitación, ya que esta incluía una regla concreta sobre la desinformación y perjuicio inminente. El Consejo revocó la decisión de Meta porque consideró que Meta no había demostrado que el contenido contribuía a ocasionar perjuicios inminentes, el estándar de las propias políticas de la compañía. El hecho de que la publicación estuviese dirigida a entidades políticas públicas y que no recomendara ningún tratamiento a los individuos, junto con el hecho de que los fármacos no estuviesen disponibles en Francia sin prescripción médica, fueron factores importantes que se tuvieron en cuenta. El Consejo recomendó a Meta la adopción de medidas menos intrusivas para abordar las publicaciones que criticaran las posturas de las autoridades sanitarias y que supusieran posibles perjuicios (no inminentes), incluidos los vínculos a sitios-web de organismos de la sanidad pública.



## Publicación de un usuario de Myanmar/Birmania acerca de los musulmanes

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de enero de 2021	Myanmar/Birmania, Francia, China	Lenguaje que incita al odio

Otra decisión estuvo relacionada con la publicación de un usuario de Myanmar/Birmania, quien compartió dos fotografías que se hicieron virales de un niño pequeño de etnia kurda que se ahogó mientras trataba de llegar a Europa en septiembre de 2015. La publicación cuestionaba la falta de respuesta de los musulmanes en general al trato que reciben los musulmanes uigures en China, en comparación con los asesinatos que siguieron a la publicación de las caricaturas del profeta Mahoma en Francia. Meta tradujo el texto que estaba escrito en birmano como “obviamente, los musulmanes tienen problemas psicológicos”, mientras que los traductores del Consejo sugirieron la alternativa “esos varones musulmanes tienen problemas de mentalidad”. Al revocar la decisión de Meta de eliminar el contenido, el Consejo determinó que, vista en contexto, la publicación era un “comentario sobre la aparente incoherencia de las reacciones de los musulmanes ante los acontecimientos de Francia y China”.

A pesar de reconocer que el uso de lenguaje que incita al odio en línea en Myanmar/Birmania ha estado vinculado a acusaciones de posibles crímenes contra la humanidad y genocidio contra el grupo minoritario musulmán Rohingya, el Consejo determinó que la publicación “no fomentaba el odio ni incitaba de manera intencionada a ninguna forma de perjuicio inminente”. Por tanto, la decisión que había tomado Meta al identificar que la publicación infringía la norma comunitaria de lenguaje que incita al odio era incorrecta.







## Armenios en Azerbaiyán

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de enero de 2021	Armenia, Azerbaiyán	Lenguaje que incita al odio

De las primeras cinco decisiones, la única que ratificó la decisión de Meta fue la que se tomó en noviembre de 2020 en relación a la publicación de fotografías históricas descritas como iglesias de la capital de Azerbaiyán, Bakú. En el texto en ruso que acompañaba a las fotografías se afirmaba que los armenios habían construido la ciudad de Bakú y que su legado, incluidas estas iglesias, se había destruido. El usuario utilizó en la publicación hashtags que pedían el fin de las agresiones y el vandalismo azerbaiyanos. La publicación utilizaba el término “taziks” para describir a los azerbaiyanos y afirmaba que “a diferencia de los armenios, carecían de historia”.

El término “taziks” significa “palangana”, aunque parece que se usó como juego de palabras que hace referencia a la palabra “aziks”, un término despectivo para los azerbaiyanos que está incluido en la lista interna de Meta de insultos. En esta decisión, se estimó que “el contexto en el que se empleaba el término dejaba claro que se usó con la intención de deshumanizar a su objetivo”. Por ello, el Consejo determinó que la eliminación de la publicación en este contexto era coherente con las responsabilidades de Meta de proteger los derechos humanos. También se estimó que la “acumulación de este tipo de contenido podía generar un entorno más proclive a los actos de discriminación y violencia”. Por tanto, se recomendó a Meta que informara al usuario de las razones por las que existía la infracción (en este caso el uso de un único término) para que este pudiera volver a publicar el resto del mensaje si así lo deseaba.

### Al anunciar la primera ronda de decisiones, el Consejo hizo la observación siguiente:

*Creemos que las primeras decisiones de casos que ha tomado el Consejo asesor de contenidos demuestran el compromiso por controlar la responsabilidad de Meta, de defender los intereses de los usuarios y las comunidades de todo el mundo y de remodelar el enfoque de Facebook hacia la moderación de contenidos. Este es el comienzo de un proceso que llevará tiempo, aunque esperamos poder compartir el progreso realizado con muchas otras decisiones de casos del Consejo.*



# Decisión relacionada con la suspensión del expresidente Trump

**FECHA DE PUBLICACIÓN**  
5 de mayo de 2021

**PAÍS**  
Estados Unidos

**NORMA COMUNITARIA**  
Personas y organizaciones peligrosas

A las 16:21 (hora estándar del este de Estados Unidos) del 6 de enero de 2021, horas después de que una multitud entrara a la fuerza en el edificio del Capitolio en Washington, D.C. y arrasara sus salones, el por entonces presidente Donald J. Trump publicó un vídeo de un minuto de duración en Facebook, así como en su cuenta de Instagram, en el que decía lo siguiente:

*Sé que os duele, que os sentís heridos. Nos han robado las elecciones. Ha sido un triunfo aplastante y todos lo saben, sobre todo los del otro bando, pero debéis ir a casa. Debemos tener paz. Debemos tener ley y orden. Debemos respetar a la maravillosa gente que respeta la ley y el orden. No queremos que nadie salga herido. Son tiempos muy difíciles.*

El vídeo continuaba:

*Nunca antes había ocurrido nada así, que nos robaran a todos: a mí, a vosotros, al país. Estas elecciones han sido un fraude, pero no podemos actuar de manera que favorezca a esas personas. Debemos tener paz. Así que volved a casa. Os queremos. Sois muy especiales. Habéis visto lo que ocurre. Habéis visto la forma tan vil y malvada con la que tratan a los demás. Sé cómo os sentís. Pero volved a casa y hacedlo en paz.*

Una hora y 20 minutos más tarde, Meta eliminó el vídeo por infringir la norma comunitaria de personas y organizaciones peligrosas. A las 18:07, mientras la policía protegía el Capitolio, el señor Trump publicó una declaración en Facebook:

*Estas son las cosas que pasan cuando una victoria aplastante y respetable en las elecciones se arrebatada de forma tan feroz y abrupta a los grandes patriotas que*



Credit: Tyler Merbler

*han recibido un trato terriblemente injusto durante tanto tiempo. Volved a casa con amor y en paz. ¡Recordad siempre este día!*

Ocho minutos más tarde, Meta eliminó esta publicación de Facebook por infringir la norma comunitaria de personas y organizaciones peligrosas. Asimismo, bloqueó la capacidad de Trump para publicar contenido en Facebook e Instagram.

Al día siguiente, el 7 de enero, tras revisar las publicaciones de Trump, sus comunicaciones más recientes fuera de Facebook e información adicional acerca de la gravedad de los actos violentos que tuvieron lugar en Washington, Meta amplió el bloqueo de las cuentas de Trump “de forma indefinida y al menos durante las siguientes dos semanas hasta que la transición de poder se haya completado de forma pacífica”.

## Remisión del caso al Consejo por parte de Meta

Dos semanas más tarde, el 21 de enero, el día siguiente a la investidura del presidente Joe Biden, Meta remitió al Consejo la decisión de suspender el acceso de Trump a sus cuentas de Facebook e Instagram de manera indefinida. La compañía también solicitó al Consejo observaciones o recomendaciones sobre suspensiones de cuentas de usuarios que son líderes políticos.

Como era de esperar, la suspensión indefinida de la cuenta de Instagram y la página de Facebook de Trump atrajo la atención del público, la política y la prensa, y generó un intenso debate acerca de la imparcialidad, la complejidad y las limitaciones de las políticas que rigen la moderación de contenidos en las plataformas de redes sociales. La cuestión también volvió a llamar la atención del Consejo acerca del “valor que aporta la supervisión independiente de las decisiones más importantes de compañías como Facebook”. El papel clave que desempeñó el Consejo para resolver esta tensión también se hizo eco en *The Washington Post*:

*“Funcionarios y legisladores de todo el mundo que buscan diseñar nuevos marcos de trabajo para regular el sector de los medios siguen con atención las acciones del Consejo. Si el Consejo emite un juicio positivo, es probable que se reduzca la urgencia por la necesidad de regulación. Si el juicio es negativo, podría intensificarse la demanda para crear barreras legales más estrictas en relación con la moderación de contenidos en numerosos países”.*<sup>5</sup>

Por su parte, *The New York Times* también abordó la cuestión.<sup>6</sup>

*Esta decisión tiene importantes consecuencias no solo para los políticos de América, sino para la regulación de las redes sociales y la posible aparición de un nuevo poder corporativo transnacional en un momento en el que parece no existir ningún poder válido.*

*En su decisión, el Consejo consideró si la suspensión, por parte de Meta, del acceso de Donald Trump para publicar contenidos en Facebook e Instagram infringía las políticas de Meta, sus valores y las normas internacionales de derechos humanos.*

El Consejo tuvo en cuenta varias fuentes de información: la lógica que siguió Meta para tomar su decisión, las respuestas de Meta a las preguntas específicas que planteó el Consejo, la declaración del creador de contenidos que un administrador de la página y el Centro Americano para la Ley y la Justicia presentaron al Consejo en nombre de Trump, junto con una investigación independiente.

## Explicación de Meta

En los casos en los que Meta debe tomar una acción de emergencia que es de interés general, a menudo comparte la lógica que ha seguido para tomar su decisión a través de una publicación en Newsroom. En este caso, Meta indica que eliminó las dos publicaciones de contenido del 6 de enero de 2021 por infringir la norma comunitaria sobre personas y organizaciones peligrosas. De manera más específica, el contenido se eliminó por infringir la política que

5 “Facebook’s New ‘Supreme Court’ Overturns Firm In First Rulings,” Elizabeth Dwoskin, Craig Timberg, *The Washington Post*, 28 de enero del 2021

6 “Trump Wants Back on Facebook. This Star-Studded Jury Might Let Him.” Ben Smith, *The New York Times*, Jan. 24, 2021



**Es posible que el juicio de Trump no satisfaga a todo el mundo. Sin embargo, este texto de 38 páginas contribuye a que nos replanteemos cómo hay que gestionar la libertad de expresión en la era del caos informativo.”**

Alan Rusbridger  
Miembro del Consejo



prohíbe “la exaltación, el apoyo o la representación de eventos designados como violentos”. Según Meta, el contenido también infringía la parte de la política que prohíbe “la exaltación de individuos que participan en actos de violencia organizada”.

Meta apunta que pese al hecho de que Trump en su vídeo pidió a las personas que volvieran a sus casas en paz, también aludió, en repetidas ocasiones, que había habido fraude en las elecciones y sugería un propósito común al decir “Sé cómo os sentís”. Dada la inestabilidad que se vivía en el momento en que hizo estos comentarios y el tono general de sus palabras, Meta concluyó que la frase “Os queremos. Sois muy especiales” tenía como objetivo exaltar a las personas que quebrantaron la ley al irrumpir en el Capitolio. La compañía también estima que la segunda publicación también exaltaba el evento, ya que Trump se refería a las personas que asaltaron el Capitolio como “grandes patriotas” e instaba a la gente a “recordar siempre este día”.

En cuanto a la posterior decisión de ampliar el bloqueo de 24 horas impuesto a Trump para publicar contenidos de manera indefinida, Meta indicó que tras realizar una evaluación adicional de la situación en evolución y de los detalles que se iban conociendo acerca de los actos violentos que tenían lugar en el Capitolio, el bloqueo de 24 horas no bastaba para abordar el “riesgo de que Trump usara su presencia en Facebook e Instagram para fomentar una escalada de la violencia”. La compañía mencionó un boletín del

Sistema Nacional de Alerta contra el Terrorismo del Departamento de Seguridad Nacional estadounidense (DHS), que describía un “entorno de amenaza elevado en los Estados Unidos que el DHS estima que persistirá durante varias semanas tras la investidura presidencial”.

Tras la investidura de Biden el 20 de enero, Meta mantuvo el bloqueo con carácter “indefinido” e indicó que el bloqueo no se anularía tras vencer ningún plazo ni con el cumplimiento de ningún criterio, por lo que su duración quedaba sometida al criterio de la compañía.

Según la compañía, la decisión de suspender la capacidad de publicación de Trump con carácter indefinido se tomó en base a las normas internacionales sobre derechos humanos, sobre todo en base al artículo 19 del Pacto Internacional de Derechos Civiles y Políticos (ICCPR) y la Observación general 34 del Comité de Derechos Humanos de la ONU, que permite aplicar restricciones necesarias y proporcionadas al derecho de libertad de expresión en situaciones de emergencia pública que supongan una amenaza para la vida de la nación.

A pesar de que Meta cuenta con una “concesión de interés periodístico” que le permite mantener en la plataforma contenido que infrinja sus políticas si considera que tal contenido es “relevante y de interés público”, Meta explicó que optó por no aplicar esa concesión a las publicaciones objeto de este caso.

### **El Consejo ratifica la decisión inicial de suspender la cuenta pero revoca la decisión de mantener la suspensión con carácter indefinido.**

El 5 de mayo, el Consejo [publicó su decisión](#) y emitió varias recomendaciones. El Consejo estimó que, dada la gravedad de las infracciones y el continuo riesgo de escalada de la violencia, consideraba justificada la medida de suspender las cuentas de Trump el 6 de enero de 2021 y de ampliar la suspensión el 7 de enero. El Consejo determinó que las dos publicaciones en cuestión infringían las normas comunitarias de Facebook e Instagram. Las normas de Facebook prohíben la publicación de contenido que exprese apoyo o exaltación de grupos, líderes o personas que participen en actividades como por ejemplo terrorismo, violencia organizada o actividades criminales.

En su decisión, el Consejo indicó lo siguiente:

*En el momento en que se hicieron las publicaciones, se estaban produciendo los actos violentos en el Capitolio. En ambas publicaciones, se exaltaba o apoyaba a las personas que participaban en actos violentos. Las palabras “Os queremos. Sois muy especiales” de la primera publicación, así como “grandes patriotas” y “recordad siempre este día” de la segunda publicación se interpretaron como acto de exaltación o apoyo a las personas implicadas en los actos violentos y los eventos que tenían lugar en el Capitolio ese día.*

Sin embargo, la decisión también estimó como inapropiada la decisión de Meta de imponer una “sanción indeterminada y sin fundamentos normativos de suspensión indefinida”. El Consejo afirma que la aplicación de una sanción de duración indefinida “carece de base” en las normas comunitarias. La decisión sostiene que la vaguedad de la sanción impuesta “infringe los principios de libertad de expresión...” También añade que “No queda claro qué normas conducirían a esta penalización o qué normas se emplearán para mantenerla o eliminarla”.

El Consejo rechazó la solicitud de Meta de “apoyar restricciones indefinidas, que se impongan y se anulen sin criterios claros”. Asimismo, insistió en que “en un plazo de seis meses tras la fecha de esta decisión, Meta debe examinar de nuevo la sanción arbitraria impuesta el 7 de enero y decantarse por una medida adecuada. Esta sanción debe basarse en la gravedad de la infracción y en la posibilidad de que se produzcan perjuicios en el futuro. Asimismo, debe ser coherente con las reglas que aplica la plataforma en casos de infracciones graves, que deben ser, a su vez, claras, necesarias y proporcionadas.”

## Respuesta de Meta a la decisión del Consejo

El 4 de junio, Meta respondió a la decisión del Consejo, también a la solicitud de revisar la sanción indefinida y decantarse por una respuesta proporcionada conforme a las normas que se aplicaban a los demás usuarios

“

**Meta tiene que saber que no puede inventarse sanciones sobre la marcha. Debe atenerse a sus propias reglas.”**

Helle Thorning-Schmidt  
Copresidente del Consejo asesor  
de contenido



de Facebook. La compañía anunció la adopción de un nuevo conjunto de protocolos de cumplimiento de normas “para casos excepcionales como este”. Según esos protocolos, Meta anunció la suspensión de la cuenta de Trump por dos años a contar desde el 7 de enero de 2021. “Dada la gravedad de las circunstancias que motivaron la suspensión de Trump, creemos que sus acciones constituyen una infracción grave de nuestras reglas y que, por tanto, merece la máxima sanción disponible según los nuevos protocolos de cumplimiento normativo”. Al aproximarse a la fecha de vencimiento de la sanción, Meta se comprometió a lo siguiente:

*Buscar expertos para evaluar si ya no existen riesgos para la seguridad pública. En la evaluación se prestará atención a diferentes factores externos, incluidos los actos de violencia, las restricciones de reunión pacífica, así como cualquier otro indicador de disturbios civiles. Si se determina que siguen existiendo riesgos graves para la seguridad pública, se ampliará la restricción por un período determinado y se seguirán sometiendo esos riesgos a nuevas evaluaciones hasta que hayan desaparecido. Cuando finalmente se levante la sanción, podrá aplicarse un conjunto de sanciones cada vez más estrictas en el caso de que el Sr. Trump cometa más infracciones en el futuro, pudiendo llegar a la eliminación permanente de las páginas y las cuentas.*

# Garantizar el respeto del derecho de libertad de expresión

De las 20 decisiones que el Consejo tomó en 2021, una gran parte exigió a Meta la restauración de contenido que se había eliminado de manera incorrecta. En estos casos, el Consejo determinó que las acciones de Meta no eran coherentes con las normas, los valores y los compromisos de la compañía con las normas internacionales sobre derechos humanos. En los cinco casos representativos que se detallan a continuación, el Consejo estimó que la decisión de Meta de restringir el derecho de libertad de expresión de los usuarios no estaba justificada dadas las circunstancias.



## Preocupación punyabí sobre la RSS en la India

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
29 de abril de 2021	La India	Personas y organizaciones peligrosas

En noviembre de 2020, un usuario compartió una publicación con vídeo en punyabí de la empresa multimedia online Global Punjab TV. En este vídeo se mostraba una entrevista de 17 minutos con un profesor a quien se describía como un activista social y defensor de la cultura punyabí. En el texto que acompañaba a la publicación, el usuario afirmaba que la organización nacionalista hindú Rashtriya Swayamsevak Sangh (RSS) y el partido que gobernaba en la India, Bharatiya Janata Party (BJP), estaban amenazando con asesinar a los sijes, una minoría religiosa de la India. La publicación aseveraba que el primer ministro Modi, anterior líder de RSS, estaba proclamando la amenaza de “genocidio contra los sijes” por consejo del presidente de la RSS.

Después de que un usuario denunciara la publicación, un moderador humano eliminó el contenido por infracción de la norma comunitaria sobre personas y organizaciones peligrosas. Ello activó una restricción automática de la cuenta del usuario. A continuación, este usuario apeló ante la compañía. Meta informó a este usuario de que no podía revisar su apelación por una reducción temporal de su capacidad de revisión a raíz del COVID-19. Tras la selección del caso por parte del Consejo, Meta restauró el contenido con retraso y admitió que su decisión inicial fue incorrecta.

Tal como indicó el Consejo en su decisión, ninguna parte del vídeo publicado infringía las políticas de Meta. El Consejo estimó que la publicación hacía hincapié en las inquietudes de esa minoría y de la oposición de la India a las que el gobierno está presuntamente discriminando. En respuesta al argumento de Meta de carecer de capacidad para revisar la apelación del usuario debido al COVID-19, el Consejo se pronunció de la misma manera que con el caso *Síntomas del cáncer de mama y desnudos*, en el que se decidió que todos los casos debían apelarse ante Meta antes de llegar al Consejo.

## Informe sobre la cronología de la implementación y respuesta de Meta a nuestras decisiones

- Según los estatutos, Meta debe implementar las decisiones del Consejo **en un plazo de siete días** tras la publicación.
- En cuanto a las 20 decisiones que se publicaron en 2021, Meta restauró o eliminó los contenidos objeto de las decisiones con arreglo al plazo establecido de siete días, excepto en los casos en los que el contenido ya se había restaurado.
- En la inmensa mayor parte de las decisiones que requerían la toma de acciones, cabe mencionar que Meta restauró o eliminó las publicaciones en el mismo día en que se publicaron las decisiones.
- Gracias al comité de implementación, actualmente compuesto por cinco miembros del consejo, el Consejo sigue exigiendo a Meta una mayor transparencia con respecto a sus procesos con el fin de identificar y emprender acciones de cumplimiento con contenidos idénticos a los contemplados en nuestras decisiones y que se presentan en contextos paralelos. Esto garantiza que las decisiones se aborden fuera de los casos específicos y que se generalicen para contenidos relevantes con contextos similares.

### Meme “dos botones”



FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
20 de mayo de 2021	Estados Unidos, Armenia, Turquía	Lenguaje que incita al odio, cruel e irrespetuoso

El 24 de diciembre de 2020, un usuario de Facebook de Estados Unidos publicó un comentario que contenía una adaptación del meme “dos botones”, del cómic “lucha diaria”, al contexto político de Turquía. El meme (publicado en 2014) muestra un personaje de animación tratando de pulsar uno de dos botones etiquetados con afirmaciones contradictorias. En la publicación, el personaje de animación cuya cara estaba cubierta con una bandera turca se encontraba debajo de dos botones rojos con afirmaciones en inglés: “The Armenian Genocide is a lie” (el genocidio armenio es una mentira) y “The Armenians were terrorists that deserved it” (los armenios eran terroristas que se lo merecían). Facebook eliminó la publicación e informó al autor después de que un moderador de contenido humano considerara que el meme infringía la norma comunitaria de lenguaje que incita al odio y de que otro indicara que también infringía la norma comunitaria de contenido cruel e irrespetuoso. El usuario apeló ante Meta y, a continuación, ante el Consejo.

Nuevamente, el contexto fue un elemento clave. Meta basó su decisión de eliminar la publicación en la afirmación “los armenios eran terroristas que se lo merecían”. Sin embargo, el Consejo estimó que, a la luz del contexto, la publicación pretendía criticar la postura negacionista del gobierno turco con respecto al genocidio armenio. Por lo tanto, la publicación podía ampararse en la excepción a la norma sobre lenguaje que incita al odio que permite compartir este tipo de lenguaje para “condenarlo o concienciar al público”. El Consejo también llegó a la conclusión de que podía aplicarse la excepción de la sátira y pidió a la compañía que incluyera información sobre esa excepción en sus normas comunitarias públicas.



## Aislamiento de Öcalan

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
8 de julio de 2021	Turquía, Estados Unidos	Personas y organizaciones peligrosas

La decisión del Consejo de julio de 2021 sacó a relucir la importancia de la revisión de las acciones de Meta por parte de una entidad independiente al descubrir una parte de una instrucción que llevaba perdida tres años.

El 25 de enero de 2021, un usuario de Instagram de Estados Unidos publicó una fotografía de Abdullah Öcalan, miembro fundador del Partido de los Trabajadores del Kurdistán (PKK).

La publicación contenía un texto en inglés que animaba a hablar sobre la necesidad de poner fin al aislamiento penitenciario de Öcalan en la isla de Imrali, en Turquía, así como sobre la falta de humanidad del régimen de aislamiento. Tanto el PKK como Öcalan están designados como “entidades peligrosas” según la política sobre personas y organizaciones peligrosas de Facebook. Esta política también prohíbe la “exaltación” o el “apoyo” a esas entidades.

Tras la eliminación de la publicación por parte de un moderador por infringir la política sobre personas y organizaciones peligrosas, el usuario apeló esa decisión. Tras el rechazo de la apelación por parte de Meta, el usuario apeló ante el Consejo asesor de contenidos.

Después de que el Consejo seleccionara el caso, salió a relucir un importante aspecto que recalca la importancia de la revisión independiente que realiza el Consejo sobre las acciones de Meta. La compañía reveló que parte de una pauta interna, elaborada en 2017, permitía hablar de las condiciones de aislamiento de personas designadas como peligrosas. Esta pauta se elaboró, en parte, como respuesta a las preocupaciones por las condiciones de encarcelamiento de Öcalan y para permitir a los usuarios debatir sobre estas cuestiones en la plataforma. Al revocar la decisión de Meta de eliminar la publicación, el Consejo también expresó su inquietud ante la falta de transparencia de Meta: “La política de eliminar de forma predeterminada el contenido que muestra ‘apoyo’ a personas designadas provocó que se ignorara este error durante un período prolongado y, al mismo tiempo, mantuvo ocultas al público excepciones clave.”

### ¿Cuáles han sido las normas comunitarias más examinadas en las decisiones del Consejo?



**Lenguaje que incita al odio**

**9 decisiones**



**Personas y organizaciones peligrosas**

**5 decisiones**



**Violencia e incitación**

**2 decisiones**





## Publicación compartida de Al Jazeera

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
14 de septiembre de 2021	Israel, Egipto	Personas y organizaciones peligrosas

El 10 de mayo de 2021, un usuario de Facebook de Egipto compartió una publicación de la página verificada de Al Jazeera en árabe que incluía una fotografía y un texto en árabe. En la imagen se veían dos hombres vestidos con ropa de camuflaje y el rostro cubierto. Ambos individuos llevaban una cinta en la cabeza con la insignia de las Brigadas de Al-Qassam.

El texto rezaba: “Los líderes de la resistencia en la sala de operaciones conjuntas ofrecen a las fuerzas de ocupación una tregua hasta las 18:00 horas para que retiren a sus soldados de la mezquita de Al-Aqsa y el barrio de Sheikh Jarrah. De lo contrario, el que avisa no es traidor. Abu Obeida, portavoz militar de las Brigadas de Al-Qassam.” El usuario compartió la publicación de Al Jazeera y añadió la expresión “Ooh” en árabe. Las Brigadas de Al-Qassam y su portavoz, Abu Obeida, se consideran peligrosos en virtud de la norma comunitaria de Facebook sobre personas y organizaciones peligrosas. Meta eliminó la publicación en virtud de esta política, pero la restauró cuando el Consejo seleccionó el caso.

Meta no pudo explicar el motivo por el que dos revisores consideraron inicialmente que el contenido infringía esta política, y señaló que no se exige a los moderadores que registren su razonamiento en el caso de decisiones relativas a contenidos individuales. La decisión del Consejo subrayaba que, aunque la política sobre personas y organizaciones peligrosas de Facebook prohíbe la “canalización de información o recursos, incluidas las comunicaciones oficiales, en nombre de una entidad”, también contiene una excepción relativa al contenido publicado con “fines periodísticos”. El contenido de este caso era una copia de un artículo periodístico publicado en Al Jazeera sobre un conflicto armado de mayo de 2021 entre fuerzas israelíes y grupos militantes palestinos en Israel y Gaza, un territorio palestino gobernado por Hamas, sin ninguna otra modificación que la adición del comentario “ooh”. Así pues, concluimos que la publicación se encontraba dentro de la excepción de fines periodísticos de Facebook.

El caso también abarcaba la equidad en el trato de los usuarios, porque esta publicación, que consistía en volver a publicar un artículo periodístico de origen legítimo, se trató de forma diferente al contenido publicado por la propia agencia de noticias. Tal como se indica en nuestra decisión, las personas tienen el mismo derecho de volver a publicar noticias que los grupos mediáticos que las publican en primer lugar. También señalamos que, mientras la página de Al Jazeera se benefició del sistema de verificación cruzada que Meta aplica a determinadas cuentas de nivel superior, el usuario que compartió el contenido, no.



## Protestas en Colombia

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
27 de septiembre de 2021	Colombia	Lenguaje que incita al odio

En junio de 2021, la página de Facebook de una cadena de noticias regional de Colombia compartió un vídeo que mostraba a unos manifestantes que caminaban detrás de una pancarta con el eslogan “SOS COLOMBIA”. El audio del vídeo contenía coros y canciones en español, en los que los manifestantes criticaban al presidente colombiano, Iván Duque, con insultos homofóbicos que provocaron que Meta eliminase el vídeo (que se había originado en Tik-Tok) por infringir su norma comunitaria de lenguaje que incita al odio. Esta norma prohíbe el contenido que “describa o señale negativamente a personas mediante insultos (...) que se usan a modo de calificaciones vejatorias”. La política prohíbe específicamente las “calificaciones inherentemente ofensivas” que hagan referencia a características protegidas, incluida la orientación sexual.

En la apelación contra la decisión de Meta de eliminar el vídeo, el usuario indicó que era un periodista local que informaba sobre las noticias de su provincia. El usuario mantuvo, además, que al compartir el vídeo, su canal no pretendía infligir ningún daño, sino simplemente mostrar a un grupo de jóvenes que protestaban pacíficamente y que pedían derechos usando su lenguaje habitual.

Al decidir que Meta debía restaurar el vídeo, el Consejo confirmó que, aunque el contenido de la publicación infringía la norma comunitaria de lenguaje que incita al odio de Facebook, ya que compartir un insulto puede contribuir a generar un entorno de intimidación y exclusión para los miembros del colectivo LGBT, Meta había cometido un error en su eliminación porque no aplicó la “concesión de interés periodístico” a la publicación. Esta disposición puede permitir que se mantenga contenido que infrinja normas en función de su nivel de interés público y riesgo de perjuicios.



# Protección de los usuarios frente a contenido perjudicial

En ciertas decisiones clave dictadas en 2021, el Consejo ratificó la decisión de Meta de retirar contenidos, ya que consideró que la limitación de la libertad de expresión del usuario puede justificarse bajo determinadas circunstancias. Estas decisiones también dejaban claro que la eliminación de contenidos de Meta puede ser coherente con las responsabilidades sobre derechos humanos de la compañía, especialmente cuando la eliminación de las manifestaciones de un usuario era necesaria y proporcionada en lo que respecta a los derechos de otras personas.



## Representación de Zwarte Piet

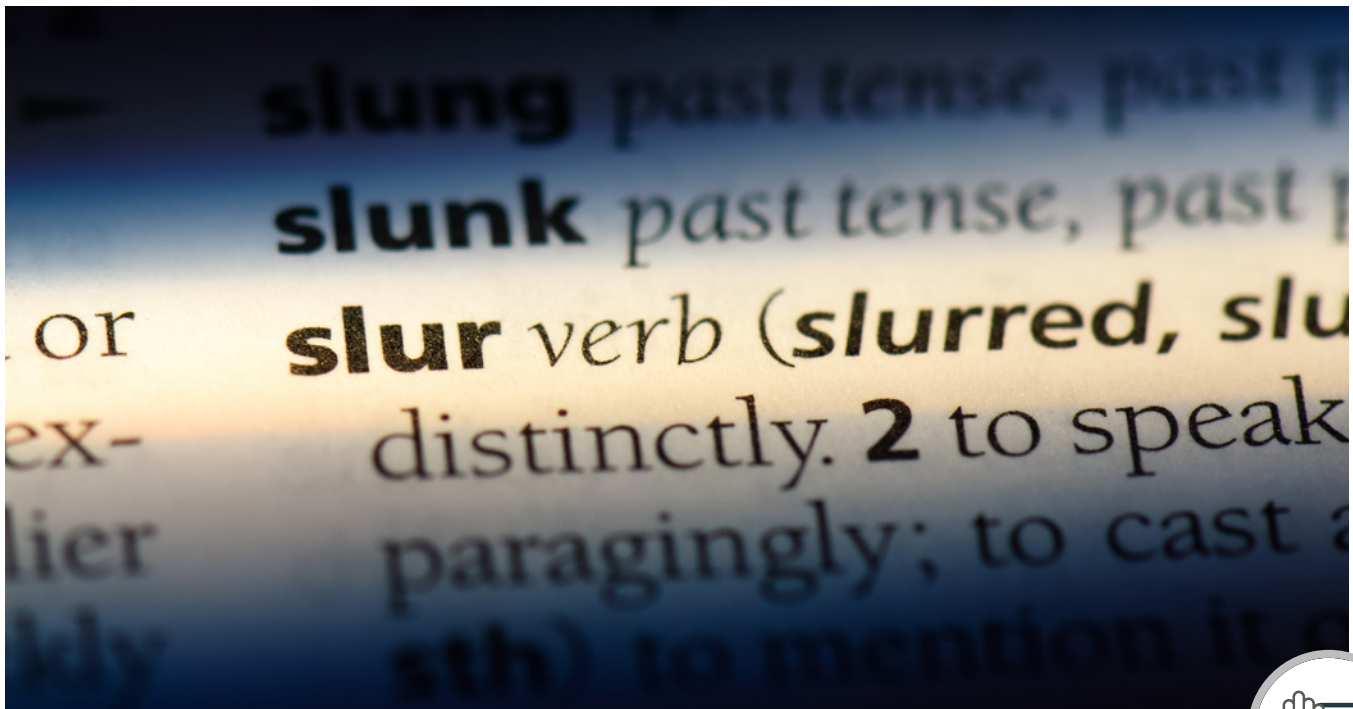
FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
13 de abril de 2021	Holanda	Lenguaje que incita al odio

El 5 de diciembre de 2020, un usuario de Facebook de los Países Bajos compartió en su biografía una publicación, que incluía un texto escrito en neerlandés y un vídeo de 17 segundos en el que aparecía un niño que se encontraba con tres adultos. Uno iba caracterizado de “Sinterklaas” o Santa Claus, mientras que los otros dos aparecían con la cara pintada de negro y llevaban pelucas afro bajo sus sombreros y prendas coloridas de estilo renacentista, representando una caricatura holandesa conocida como “Zwarte Piet”. En español, el nombre se podría traducir como “Pedro el Negro”.

Apoyamos la decisión de Meta de retirar la publicación por infringir su norma comunitaria de lenguaje que incita al odio. La mayoría de los miembros del Consejo opinó que el contenido contenía “caricaturas estrechamente vinculadas con estereotipos negativos y racistas, que, conforme al criterio de parte de la sociedad neerlandesa, sustentan el racismo sistémico en los Países Bajos”. Para llegar a esta decisión, la mayoría de los miembros del Consejo pudo examinar estudios que indicaban que “durante la fiesta de Sinterklaas, los niños negros estaban asustados, se sentían inseguros en sus hogares y temían asistir al colegio”.

Aunque una minoría de los miembros del Consejo no percibió suficiente evidencia para relacionar directamente la eliminación de este fragmento de contenido con una supuesta reducción del perjuicio, la mayoría determinó que permitir la prodigalidad de este tipo de contenidos en la plataforma contribuiría configurar un entorno discriminatorio para las personas negras, que resultaría degradante y hostigador. La mayoría observó que:

*A escala, la política [de Meta] es clara y garantiza la dignidad, la seguridad y la opinión de las personas negras en la plataforma. Para alcanzar este objetivo, es aceptable restringir la expresión de las personas que comparten representaciones de caras pintadas de negro si su objetivo no es el de reprobear el racismo.*



## Insultos en Sudáfrica

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
28 de septiembre de 2021	Sudáfrica	Lenguaje que incita al odio

En mayo de 2021, un usuario de Facebook hizo una publicación en inglés en un grupo público cuyo objetivo, según su descripción, era promover la “apertura de mente”. La publicación hablaba sobre el “multirracismo” en Sudáfrica y señalaba que la pobreza, el sinhogarismo y la cantidad de personas sin tierra han aumentado entre la población negra del país desde 1994. El usuario indicaba que las personas blancas poseen y controlan la mayor parte de la riqueza, y que las personas negras ricas quizá son propietarias de algunas empresas, pero no tienen el control de estas. La publicación indicaba que las personas con una visión contraria “have their heads examined” (deberían hacérselo mirar), describiendo un grupo de personas negras que usaban insultos raciales y que Meta explicó que infringían su política de lenguaje que incita al odio.

Ratificamos la decisión de Meta de eliminar este contenido, ya que encontramos que uno de los términos usados era una “palabra dañina que transmite odio en Sudáfrica”. Subrayamos que la discusión abierta sobre la desigualdad socioeconómica es clave para avanzar en la igualdad. A pesar de ello, consideramos que el usuario podía haber entablado esta discusión y apelar a las emociones de su audiencia sin hacer referencia a ese insulto. Además, un insulto en particular tiene una conexión específica con la historia del apartheid del país, y su uso todavía se asocia con la degradación y exclusión de las personas negras. La decisión reiteraba la importancia de que Meta proporcionara a las personas cuyo contenido eliminaba, bajo la política de lenguaje que incita al odio, una explicación que especificara la regla infringida.



## Presunto delito en Raya Kobo

FECHA DE PUBLICACIÓN	PAÍS	NORMA COMUNITARIA
14 de diciembre de 2021	Etiopía	Lenguaje que incita al odio

Una decisión de diciembre de 2021 consideró que una publicación que incluía rumores no verificables sobre un grupo étnico infringía las normas comunitarias de Meta y su eliminación era coherente con los principios sobre derechos humanos.

A finales de julio de 2021, un usuario de Facebook hizo una publicación en su biografía, en el idioma etíope amárico, alegando que el Frente de Liberación Popular de Tigray (TPLF) había asesinado y violado a mujeres y niños, además de saquear propiedades en el distrito etíope de Raya Kobo y ciudades cercanas de la región Amhara. En la publicación, el usuario afirmaba que civiles étnicos tigrayanos habían ayudado al TPLF a cometer estas atrocidades. La publicación concluía con las siguientes palabras: “Nuestra lucha nos liberará”. Además, el contenido, que había sido visto por miles de hablantes de amárico en las 24 horas que permaneció en línea, incluía comentarios de algunos de ellos con llamadas a la venganza.

Un moderador de contenidos del equipo de revisión de contenido en amárico de Meta determinó que la publicación infringía la norma comunitaria de lenguaje que incita al odio y la eliminó. La política prohíbe el “lenguaje violento” dirigido a las personas por su raza, etnia o nacionalidad. El usuario apeló la decisión ante Meta, que, después de una segunda revisión por otro moderador del equipo de revisión de contenido en amárico, confirmó que la publicación infringía las políticas de Facebook.

Después de que el Consejo seleccionara el caso para revisión, Meta identificó la eliminación como un “error del sistema de cumplimiento de normas” y restauró el contenido. Meta basó esta decisión en la evaluación de que las declaraciones de la publicación no llegaban al nivel de lenguaje que incita al odio.

Discrepamos ya que a la explicación de Meta le “faltaban detalles y era incorrecta”. Aplicamos la norma comunitaria sobre violencia e incitación a esta publicación, y consideramos que infringía la prohibición de Facebook sobre los “rumores de desinformación y no verificables que contribuyen al riesgo de violencia inminente o daño físico”. El Consejo determinó que los “rumores que aseguraban la complicidad de un grupo étnico en atrocidades masivas eran peligrosos y aumentaban significativamente el riesgo de violencia inminente durante un conflicto violento” como el que ocurre actualmente en Etiopía.

# Normas internacionales sobre derechos humanos en la toma de decisiones del Consejo



Parte de nuestra responsabilidad es garantizar que el enfoque de Meta sobre la moderación de contenidos se base en el respeto a los derechos humanos. Así se refleja en nuestra Acta constitutiva y en los [compromisos de derechos humanos corporativos](#) de Meta. Por tanto, no solo determinamos si las decisiones de contenido de Meta siguen las reglas y los valores para contenidos de la compañía, sino también examinamos si se ajustan a las normas internacionales sobre derechos humanos.

Las normas internacionales sobre derechos humanos proporcionan un marco universal para proteger la libertad de expresión y otros derechos humanos. También ofrecen un marco coherente y global en el que Meta debe rendir cuentas de su impacto en todo el mundo. La política de derechos humanos corporativa



**Es hora de hablar sobre cómo debemos crear tecnología que sea intrínsecamente respetuosa con los derechos humanos”.**

Julie Owono  
*Miembro del consejo*



de Meta establece sus compromisos con respecto a los derechos humanos, incluida la forma en que desarrolla sus políticas sobre contenidos. La política también determina que los Principios Rectores sobre las Empresas y los Derechos Humanos (PRNU) de la ONU fundamentan la creación del Consejo.

Creemos que los derechos humanos deben ser la base de nuestro trabajo y, si hacemos que Meta los respete, Facebook e Instagram serán lugares mejores para los usuarios. Con el tiempo esperamos que, gracias a nuestras decisiones, el enfoque de Meta sobre la moderación de contenidos se alinee mejor con las normas internacionales sobre derechos humanos.

## Un nuevo reto

Que una corporación como Meta aplique normas internacionales sobre derechos humanos a la moderación de contenidos es un reto relativamente nuevo. Ninguna otra entidad está involucrada en el tipo de trabajo que actualmente desempeña el Consejo.

El uso de un marco legal creado por estados para estados, a fin de interpretar las responsabilidades sobre derechos humanos de una compañía de redes sociales, plantea muchas preguntas. ¿Hasta qué punto se debe someter la moderación de contenidos de Meta a las mismas estructuras aplicadas a la regulación del lenguaje de los gobiernos en otros lugares? ¿Cómo debe considerar este enfoque los retos de moderación

de contenidos a escala? ¿Cómo debemos explorar los daños que ocasionen las elecciones de diseño de Meta, incluida la amplificación algorítmica de contenido? Mediante nuestras decisiones hemos explorado cómo se pueden usar las normas internacionales sobre derechos humanos para definir las responsabilidades sobre derechos humanos de Meta, ya que pueden diferir de las que son aplicables a los estados. En nuestras decisiones de *Representación de Zwarte Piet y Armenios en Azerbaiyán*, por ejemplo, ratificamos la decisión de Meta de eliminar el contenido, a la vez que reconocíamos que, probablemente, las normas internacionales sobre derechos humanos no permitirían que un estado castigara el mismo lenguaje con medidas penales o administrativas.

## Construcción de una organización basada en los derechos humanos

Como organismo cuyas decisiones afectan a personas de todo el mundo, es vital que respetemos los derechos humanos en todos los aspectos de nuestro

trabajo. El Acta constitutiva del Consejo requiere que se preste especial atención a los derechos humanos al tomar decisiones. Queremos seguir los PRNU, por lo que nuestro trabajo es legítimo, accesible, equitativo, transparente, compatible con los derechos, una fuente de aprendizaje continuo y está basado en la participación y el diálogo. En 2020, el Consejo encargó un informe de seguimiento a la organización sin ánimo de lucro Business for Social Responsibility (BSR) que indicó que el Consejo había “progresado correctamente” en las recomendaciones que había hecho en el informe anterior.

## Cómo los derechos humanos sustentan nuestro enfoque de libertad de expresión

Entre las fuentes de autoridad que guían las decisiones del Consejo son las normas internacionales sobre derechos humanos, con el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos entre los más citados. El ICCPR es un tratado de derechos

## Reacción de las organizaciones al trabajo del Consejo

Los académicos, la sociedad civil y las instituciones nacionales e internacionales han empezado a analizar el trabajo y las decisiones del Consejo. Los puntos destacados incluyen:

- En diciembre de 2020, el Relator especial de la ONU sobre cuestiones de minorías, Fernand de Varennes, nos urgió a considerar los derechos de las minorías, y dijo del Consejo que era “una iniciativa innovadora y ambiciosa para regular la expresión en línea, en particular, lenguaje que incita al odio, esencial para la protección efectiva de las minorías vulnerables de todo el mundo”.
- En marzo de 2021, un analista de investigación legal de la Biblioteca de Derecho del Congreso examinó cómo las cinco primeras decisiones del Consejo aplicaban las leyes internacionales de derechos humanos a las compañías de redes sociales.
- En septiembre de 2021, las organizaciones, incluidas Access Now, artículo 19 y Tamleh, aplaudieron la decisión de compartir la publicación de Al Jazeera que tomó el Consejo como un “paso en la dirección correcta” y dijeron que era un caso “emblemático de la arbitrariedad sistemática y la sobreimposición no transparente de Facebook” de la norma comunitaria sobre personas y organizaciones peligrosas.
- En diciembre de 2021, la Comisión de Estados Unidos para la Libertad Religiosa Internacional mencionó nuestro trabajo al aplicar normas de derechos humanos a la moderación de contenido.
- “Lawfare” creó un blog sobre el Consejo asesor de contenido que compila los comentarios y el análisis de nuestras decisiones hasta ahora.
- Nuestras decisiones también se trasladan a las bases de datos legales internacionales Westlaw y Lexis/Lexis+.

humanos global que Meta cita en su política de derechos humanos corporativa, por la que se compromete voluntariamente a respetar los derechos humanos que describe. El trabajo del Relator especial de la ONU sobre la libertad de opinión y expresión, en particular un [informe sobre moderación de contenido de 2018](#) que estudió cómo las compañías privadas pueden respetar mejor los derechos humanos, ha sido especialmente valioso para el Consejo.

El artículo 19, párrafo 2, del ICCPR indica que “Toda persona tiene derecho a la libertad de expresión; este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección”. Sin embargo, en línea con las normas internacionales sobre derechos humanos, el Consejo considera que el derecho de libertad de expresión no es absoluto. Aunque el derecho sea importante, se puede restringir en determinadas circunstancias. En el artículo 19, párrafo 3 del ICCPR se proporciona una prueba en tres partes para evaluar una restricción de la libertad de expresión, que toma en cuenta la legalidad, el objetivo legítimo y la necesidad y proporcionalidad. Estos principios son similares, aunque no idénticos, a las normas de libertad de expresión que se encuentran en las leyes nacionales y constitucionales de muchas naciones en las que Meta opera, incluida la Primera Enmienda de la Constitución de Estados Unidos. También tenemos en cuenta las restricciones de expresión requeridas por las leyes internacionales de derechos humanos cuando son relevantes, incluida la prohibición de la apología del odio que constituye una incitación a la hostilidad, discriminación o violencia (Artículo 20, párrafo 2 del ICCPR).

En nuestro análisis de la *legalidad* en las decisiones, examinamos si las reglas de Meta son claras para las personas que usan Facebook e Instagram. Las reglas deben ser fáciles de encontrar y comprender por los usuarios. En numerosas decisiones, hemos presionado a Meta para que publicara fragmentos de sus reglas que están ocultas al público, incluida la lista de personas y organizaciones que la compañía designa como “peligrosas”.

En nuestro análisis de *objetivo legítimo* en las decisiones, miramos si las reglas de Meta en Facebook e Instagram persiguen objetivos que se consideran motivos legítimos para restringir la expresión según las leyes internacionales de derechos humanos. La mayoría de nuestras decisiones identifican el respeto por los “derechos de los demás” como fin legítimo de las políticas de contenido de Meta que restringen el lenguaje.

En nuestro análisis de la *necesidad y proporcionalidad*, examinamos si la acción de Meta en casos determinados es necesaria para lograr el objetivo legítimo y si las medidas impuestas, como la eliminación de contenidos, es el medio menos restrictivo para lograr ese objetivo.

En las 20 primeras decisiones que tomamos expresamos nuestra preocupación, ya que varias reglas de contenido de Meta son demasiado vagas, demasiado amplias o ambiguas; por lo tanto, hemos recomendado su aclaración o que se haga pública la guía sobre la interpretación interna de estas reglas. El Consejo ha dado una importancia especial al contenido político y al contenido que conciencia sobre las violaciones de los derechos humanos, incluida la sátira política, así como la expresión artística y el debate sobre problemas de salud. También hemos cuestionado si la eliminación de contenidos es siempre una respuesta proporcionada al contenido que puede estar relacionado con un daño, pero que no causa directamente un daño inminente. Por ejemplo, en dos decisiones relacionadas con el COVID-19 (*Solicitud de tratamiento por COVID* y *Confinamiento por COVID en Brasil*) nos preguntamos si otras medidas que no llegaran a la eliminación de las publicaciones (como son pantallas de advertencia, etiquetado o disminución del rango del contenido) podrían mitigar el riesgo y proteger la libertad de expresión a la vez.

Más allá de la prueba de tres partes anterior, también hemos integrado principios sobre la igualdad y la discriminación en nuestra toma de decisiones, ya que reconocemos que personas distintas en situaciones diferentes pueden verse afectadas de manera distinta por el enfoque que tiene Meta de la moderación de contenido. Ello fundamentó recomendaciones para que Meta evaluara su impacto sobre los derechos humanos en áreas afectadas por conflictos, en idiomas



como el árabe y el amárico, así como pedir mejoras en la moderación de los birmanos. La atención del Consejo a los problemas de igualdad y de no discriminación en la decisión sobre los *síntomas de cáncer de mama y desnudos* también mostró cómo la dependencia que tiene Meta de la automatización, que llevó a la eliminación errónea de información sobre la salud específica para mujeres, puede tener impactos desproporcionados y discriminatorios en los usuarios. Por motivos similares, debemos prestar atención a la importancia de la no discriminación según el punto de vista político.

## Derechos frente a reglas: protestas en apoyo a Navalny en Rusia

En la mayoría de nuestras decisiones, el análisis de los derechos humanos ha respaldado las conclusiones de la interpretación de las reglas de Meta. Sin embargo, en un hito para el Consejo, la decisión sobre las *protestas en apoyo de Navalny en Rusia* fue el primer caso en el que anulamos la decisión de Meta basada en sus responsabilidades sobre derechos humanos, aunque la eliminación era coherente con las reglas de Meta.

El 24 de enero de 2021, un usuario en Rusia publicó imágenes, un vídeo y un fragmento de texto sobre las protestas del día anterior en apoyo al líder de la oposición Alexei Navalny en San Petersburgo, Moscú y otras comunidades del país. Otro usuario (el “Crítico de la protesta”) diciendo que, aunque no sabía qué había sucedido en San Petersburgo, todos los manifestantes en Moscú eran niños pequeños mentalmente “slow” (lentos), a los que se estaba “shamelessly used” (usando descaradamente). Otro usuario, que decía haber participado en la protesta de San Petersburgo, llamó al Crítico de la protesta “cowardly bot” (bot cobarde). El “Crítico de la protesta” denunció el comentario del Manifestante a Facebook por *bullying* y acoso.

Meta retiró la publicación por infringir su norma sobre *bullying* y acoso, que establece que una persona adulta no pública puede solicitar la eliminación de publicaciones que contengan “comentarios negativos relacionados con su carácter”. Al decidir en favor de la solicitud del manifestante de que se restaurara la publicación, el Consejo observó:

*Que este caso pone en evidencia la tensión entre las políticas que protegen a las personas frente al bullying y el acoso y la necesidad de proteger la libertad de expresión. Ello es especialmente relevante en el contexto de protestas políticas en un país en el que existen quejas creíbles sobre la ausencia de mecanismos eficaces para proteger los derechos humanos.*

Nuestra decisión determinó que, si bien la eliminación del contenido fue coherente con la norma comunitaria sobre *bullying* y acoso de Meta, se trataba de una restricción innecesaria y desproporcionada de la libertad de expresión bajo las normas internacionales de derechos humanos.

Además, el Consejo determinó, en favor del derecho de expresión del usuario, que aunque la decisión original de Meta de eliminar el contenido “puede ser coherente con una aplicación estricta de las normas comunitarias”, esta aplicación “no había considerado el contexto general y había restringido la libertad de expresión de forma desproporcionada”.

En este caso se demostró que cuando surge un conflicto entre las políticas sobre contenidos de Meta y sus responsabilidades sobre derechos humanos, **el Consejo está preparado para priorizar los derechos humanos.**

## Estudios de casos sobre normas específicas

En las 20 primeras decisiones, el Consejo consideró un amplio elenco de normas internacionales sobre derechos humanos. Todas las decisiones publicadas en 2021 hacían referencia al artículo 19 de ICCPR, los Principios Rectores sobre las Empresas y los Derechos Humanos de la ONU y la Observación general 34 del Comité de Derechos Humanos de la ONU sobre las libertades de opinión y expresión. La aplicación de estas y de otras normas internacionales sobre derechos humanos en nuestras decisiones ha aumentado nuestro conocimiento y análisis de los problemas subyacentes que dirigen muchas decisiones de moderación de contenidos. También nos ha ayudado a llegar a conclusiones más fundamentadas. Aunque cada decisión hace referencia a varias normas de derechos humanos,

proporcionamos unos ejemplos de las que resultaron ser una contribución esencial en la toma de decisiones.

## Principios Rectores sobre las Empresas y los Derechos Humanos de la ONU

En una decisión sobre la eliminación por parte de Meta de una publicación que mostraba un cinturón con abalorios, una forma de arte indígena de Norteamérica, con el título “Kill the Indian/Save the Man” (Matad al indio/salvad al hombre), pedimos a Meta que cumpliera sus responsabilidades sobre derechos humanos con las comunidades marginadas según el Principio 17 de los PRNU. El contexto cultural de la frase se proporcionó en un comentario del público de la Asociación de Asuntos Indígenas Americanos, que identificó al autor como Richard Henry Pratt, el fundador del primer internado federal para pueblos indígenas de los Estados Unidos. En contexto, la frase constituía un claro ejemplo de “respuesta a contenido ofensivo”, una forma de expresión en la que se hace referencia deliberadamente a lenguaje que incita al odio con el fin de resistir a la opresión y discriminación, no para apoyarlas o promoverlas.

En respuesta a este caso, solicitamos a Meta que cumpliera con sus responsabilidades sobre derechos humanos con las comunidades marginadas según el Principio 17 de los PRNU. Para hacerlo, Meta debe adoptar la diligencia debida sobre los derechos humanos para garantizar que sus sistemas operen equitativamente y no exacerben la opresión histórica y en curso. Ello incluye la identificación, prevención y mitigación de los impactos negativos de la moderación de contenidos en la expresión de pueblos indígenas para combatir la discriminación. Por último, indicamos nuestra expectativa de que Meta debe dar prioridad a los riesgos relacionados con grupos marginados y demostrar evidencia de mejoras continuas.

### Observación general 34

En el caso *Bot de Myanmar*, que se centraba en contenido publicado después de un golpe de estado no democrático en el país, aclaramos a Meta que se debe distinguir entre una crítica que un usuario hace del

gobierno y el lenguaje discriminatorio contra personas en función de su nacionalidad. Al hacerlo, el Consejo se basó en la Observación general 34 del Comité de Derechos Humanos de la ONU, que establece instrucciones para interpretar el artículo 19 de ICCPR, que protege la libertad de expresión globalmente. La decisión del bot de Myanmar hace referencia específicamente a los fragmentos de la Observación general que tratan sobre la importancia del lenguaje político y, en particular, del valor de permitir el debate desinhibido sobre las instituciones públicas. Aunque la libertad de expresión pueda limitarse, en determinadas circunstancias, para proteger a los individuos frente al lenguaje que incita al odio discriminatorio y que ocasiona daños, las personas deben poder decir con libertad lo que quieran sobre los estados nacionales y las instituciones públicas.

### Derecho a reparación (artículo 2 de ICCPR)

En la decisión sobre el *aislamiento de Öcalan*, se citó la declaración sobre la libertad de expresión de la Relatora Especial de la ONU que señalaba que el proceso de reparación “debería incluir un procedimiento transparente y accesible para apelar las decisiones de la plataforma y que las empresas deberían proporcionar una respuesta fundamentada que también sea de acceso público”.

Sin embargo, en este caso, las acciones de Meta se quedaron cortas en relación a estas normas. Inicialmente, la compañía informó al usuario que, debido al COVID-19, no estaba disponible el proceso de apelación. También expresamos nuestra preocupación de “que haya habido una cantidad significativa de eliminaciones que no deberían haberse llevado a cabo” porque Meta perdió un fragmento de la pauta normativa durante tres años, que claramente permitía el contenido del usuario. La compañía reclamó que no era técnicamente posible determinar cuántos fragmentos de contenido se habían eliminado cuando los revisores no disponían de la pauta normativa. Por estas inquietudes, además de otras, criticamos a Meta porque “no está respetando el derecho a reparación, lo que infringe su Política corporativa sobre derechos humanos”.

## Una tarea difícil, pero valiosa

Reconocemos que aplicar las normas internacionales sobre derechos humanos para interpretar las responsabilidades sobre derechos humanos de una compañía de redes sociales es un ejercicio nuevo y desafiante.

Sin embargo, aunque sea una tarea difícil, es valiosa. Aplicar los derechos humanos internacionales a la moderación de contenidos nos puede ayudar a garantizar que Meta respete la libertad de expresión

junto con otros derechos humanos, y a proporcionar un marco coherente y global, en el que las compañías de redes sociales rindan cuentas.

Crear un corpus de obra que muestre lo que es la moderación de contenido basada en los derechos es fundamental para el Consejo. En 2022, recurriremos al valor preferencial de nuestras decisiones anteriores mientras continuamos garantizando el respeto por los derechos humanos de las personas que usan las plataformas de Meta.

# Preguntas formuladas por el Consejo a Meta como parte de las decisiones

Como ayuda en la toma de decisiones y para urgir a Meta a que sea lo más transparente posible, enviamos preguntas a la compañía antes de deliberar sobre los casos. Al formular preguntas específicas e incluir los detalles en nuestra decisión final, podemos pedir responsabilidades a Meta y ofrecer a los investigadores nueva información sobre cómo funciona la compañía.

En el caso de las 20 decisiones que publicamos en 2021, nuestras preguntas abarcaban muchas áreas diferentes. En muchos casos, nuestras preguntas exploraban asuntos específicos que indicaron los usuarios en su declaración ante el Consejo, así como problemas que surgieron en comentarios del público. Muchas de nuestras preguntas estaban relacionadas con el trato de Meta con respecto al usuario que apeló el contenido, como por ejemplo el motivo que le dio para eliminar el contenido y si la compañía revisó su apelación original. En otros casos, solicitamos más detalles sobre las reglas de Meta, los sistemas automatizados y los equipos regionales.

Como parte de nuestras 20 primeras decisiones, formulamos 313 preguntas a Meta. De ellas, Meta respondió a 270, parcialmente a 24 y no respondió a 19. Presentamos estas cifras en interés de la transparencia y, en ningún caso, ello pretende ser una queja sobre los motivos de Meta para no responder.

Durante nuestro primer año, Meta solía ser receptiva a las consultas del Consejo y respondía a una amplia mayoría (un 86%) de nuestras preguntas. Cuando Meta no respondía a las preguntas, lo incluíamos en nuestros informes trimestrales sobre transparencia y en algunas decisiones. Más adelante en esta sección, se incluye una descripción detallada de las preguntas a las que Meta no respondió y los motivos que dio.

Después de recibir algunas respuestas tardías a nuestras preguntas de los primeros casos, trabajamos con Meta para acordar un sistema por niveles en el

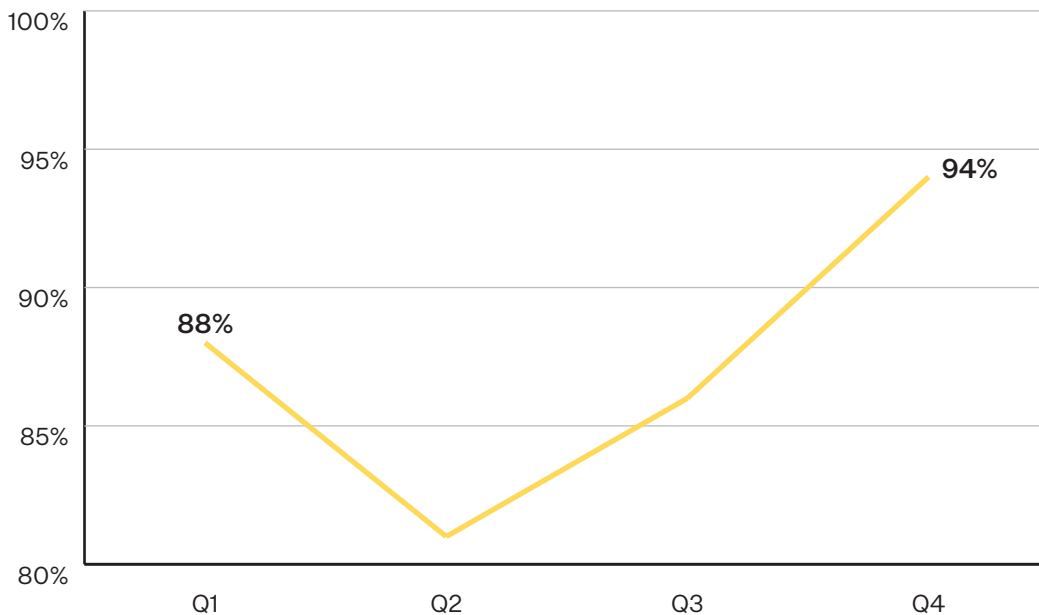
## Respuesta de Meta a las preguntas del consejo asesor de contenido



que el plazo de la compañía depende del número de preguntas que el Consejo formula. Apreciamos también que Meta estuviera dispuesta a ofrecer informes sobre problemas relacionados con sus reglas, incluido el sistema de verificación cruzada.

El porcentaje de preguntas que Meta respondió para las decisiones publicadas entre el primer y el tercer trimestre de 2021 empezó con un 88% de preguntas respondidas en el primer trimestre, en comparación con el 81% del segundo trimestre y el 86% del tercer trimestre. Sin embargo, en el caso de las decisiones publicadas en el cuarto trimestre de 2021, aumentó al 94%, ya que Meta respondió a 51 de las 54 preguntas y, además, respondió parcialmente a tres preguntas. El cuarto trimestre fue el primer trimestre en el que ninguna de las respuestas de Meta se incluyó en la categoría “sin respuesta”. Este progreso es alentador y esperamos que esta tendencia continúe durante 2022.

## El porcentaje de preguntas que Meta respondió aumentó para las decisiones publicadas en el cuarto trimestre de 2021



### Detalles sobre las preguntas a las que Meta no respondió

Meta no respondió a 19 de las 313 preguntas que formulamos como parte de nuestras decisiones publicadas en 2021. A continuación, proporcionamos más detalles sobre esas preguntas, incluido el caso con el que están relacionadas, lo que se preguntaba y el motivo que Meta dio para no responder.

- Meta no respondió a **dos preguntas** del caso *Publicación de Myanmar sobre musulmanes*. La primera pregunta se refería a si Meta había aplicado previamente la infracción de lenguaje que incita al odio al usuario y grupo en cuestión. Meta respondió que la información sobre el comportamiento anterior del usuario en la plataforma era irrelevante para la determinación del Consejo. La segunda pregunta solicitaba las traducciones al inglés del equipo de Facebook Market de los comentarios realizados en la publicación. Meta respondió que no traduciría automáticamente los comentarios como parte de su proceso de revisión de contenido y que no disponía de traducciones manuales de los comentarios.
- Meta no respondió a **una pregunta** en el caso *Armenios en Azerbaiyán*. El Consejo preguntó por qué se había eliminado una de las publicaciones

anteriores del usuario. Meta respondió que la información sobre el comportamiento anterior del usuario en la plataforma era irrelevante para la determinación del Consejo.

- Meta no respondió a **una pregunta** en el caso *Cita nazi*. El Consejo preguntó si el usuario contaba con un registro de publicaciones prohibidas o suspendidas. Meta respondió que la información sobre el comportamiento anterior del usuario en la plataforma era irrelevante para la determinación del Consejo.
- Meta no respondió a **una pregunta** en el caso *Protesta en India contra Francia*. La pregunta del Consejo solicitaba si Meta había aplicado infracciones según la norma comunitaria de violencia e incitación al usuario o grupo. Meta respondió que la información sobre el comportamiento anterior del usuario en la plataforma era irrelevante para la determinación del Consejo.
- Meta no respondió a **dos preguntas** en el caso *Preocupación punyabí sobre RSS en la India*. En la primera pregunta se pedía el idioma específico del contenido que provocó que Meta lo eliminase según la norma comunitaria de personas y organizaciones peligrosas. Meta respondió que no podía identificar el idioma específico que dio lugar a la conclusión errónea de que el contenido infringía la política de

personas y organizaciones peligrosas. La segunda pregunta incluía la petición del número de “faltas” que los usuarios necesitan para que Meta imponga una restricción de la cuenta y cuántas infracciones de la política de personas y organizaciones peligrosas eran necesarias para las restricciones de nivel de cuenta. Meta respondió que esta información no era razonablemente necesaria para la toma de decisiones de acuerdo con la intención del Acta constitutiva.

- Meta no respondió a **siete preguntas** en el caso *Suspensión del expresidente Trump*. Las preguntas que Meta se negó a contestar tenían que ver con cómo la sección de noticias y otras funciones de Facebook afectaron la visibilidad del contenido del Sr. Trump; si Meta hizo, o tiene previsto hacer, una investigación respecto de esas decisiones de contenido en relación con los acontecimientos ocurridos el 6 de enero de 2021 y se solicitó información sobre el contenido infractor de los seguidores de las cuentas del Sr. Trump. El Consejo también formuló preguntas relativas a la suspensión de otras figuras políticas y la eliminación de otros contenidos; si Meta había tenido contacto con funcionarios políticos electos u otro personal respecto de la suspensión de las cuentas del Sr. Trump; y si la suspensión o eliminación de cuentas afecta a la capacidad de los anunciantes de dirigir anuncios a las cuentas de seguidores. Meta determinó que esta información no era necesaria de manera razonable para la toma de decisiones, de conformidad con la intención del Acta constitutiva; no era posible proporcionarla desde el punto de vista técnico; estaba amparada por el privilegio entre abogado y cliente; y no podría ni debería proporcionarse por cuestiones relativas a legalidad, privacidad, seguridad o protección de datos.
- Meta no respondió a **tres preguntas** en el caso *Aislamiento de Öcalan*. En la primera pregunta se pedía si Meta podía determinar cuántos fragmentos de contenido se habían retirado incorrectamente

mientras no se pudiera aplicar la política. En la segunda pregunta se solicitaba saber cuánto contenido que mencionaba a Öcalan había eliminado Meta o dejado en los últimos cinco años. En ambas preguntas, Meta respondió que no era técnicamente posible determinar esta información. En la tercera pregunta se solicitaba saber, en parte, por qué los moderadores consideraban que la publicación era una incitación a respaldar a Öcalan y al Partido de los Trabajadores de Kurdistan (PKK). Meta respondió que, como la compañía no requería que sus revisores de contenido a escala documentasen su razonamiento para cada decisión de contenido, no podía suministrar una descripción detallada de los fundamentos de la eliminación de la publicación.

- Meta no respondió a **dos preguntas** en el caso *Insultos en Sudáfrica*. La primera pregunta solicitaba a Meta que proporcionase métricas sobre tres términos ofensivos usados en la publicación. Meta respondió que no era técnicamente posible proporcionar la información solicitada. En la segunda pregunta se solicitaba saber si en el equipo de políticas públicas de Meta en Sudáfrica había personas nativas. Meta declinó proporcionar la información solicitada, citando los estatutos del Consejo.

De las 19 preguntas que Meta no respondió, dos terceras partes (13) estaban relacionadas con casos en los que el contenido se había eliminado según la norma comunitaria de personas y organizaciones peligrosas. Eran los casos *Cita nazi*, *Preocupación punyabí sobre RSS en India*, *Suspensión del expresidente Trump* y *Aislamiento de Öcalan*. Cinco de las preguntas restantes estaban relacionadas con casos en los que el contenido se había eliminado por la política de lenguaje que incita al odio, mientras que una pregunta estaba relacionada con contenido eliminado por la política sobre violencia e incitación de Meta.

# Comentarios del público

Como parte del proceso de decisión, las personas y las organizaciones pueden presentar comentarios del público. Durante nuestro primer año, nuestras decisiones dieron voz a la población, proporcionándonos así un conocimiento crucial sobre el lenguaje, la cultura, la política y los derechos humanos. Esta información es fundamental para conseguir nuestro objetivo de mejorar la forma en que Meta trata a las personas y comunidades en todo el mundo. En numerosas ocasiones, los comentarios del público han determinado nuestras decisiones y nuestras recomendaciones a Meta.

En cuanto a las 20 decisiones publicadas en 2021, recibimos 9986 comentarios del público procedentes de personas y organizaciones de todo el mundo. La inmensa mayoría de ellos (el 97%) se presentaron debido a nuestra decisión sobre la suspensión del *expresidente Trump*, pero es importante subrayar la calidad y el valor de los

conocimientos que recibimos en muchos casos, incluso cuando el número absoluto de comentarios pudiera ser limitado. Algunos problemas son profundamente complejos y requieren conocimientos más especializados de lo que los comentarios del público pueden ofrecer.

El 96% de los comentarios del público provenían de Estados Unidos y Canadá, en gran medida debido a los comentarios de Estados Unidos sobre la decisión de la *suspensión del expresidente Trump*. El 95% de los comentarios del público fueron de personas individuales, mientras que el 5% lo fueron de organizaciones.

Un tercio de los comentarios del público se hicieron de forma anónima. Es interesante que se utilizase tanto esta opción que permite compartir conocimiento, pero proteger el derecho a la privacidad. Las tres decisiones en las que más de la

## **Decisiones del Consejo asesor de contenido que han recibido más comentarios del público en 2021:**

1. Suspensión del expresidente Trump
2. Armenios en Azerbaiyán
3. Confinamientos por COVID en Brasil
4. Publicación compartida de Al Jazeera
5. Síntomas del cáncer de mama y desnudos



mitad de los comentarios publicados se presentaron anónimamente (*Preocupación punyabí sobre la RSS en India*, *Meme “dos botones”* y *Protestas en Colombia*) estaban relacionadas con contenido que criticaba a gobiernos actualmente en el poder.



**Gracias a los comentarios del público, el Consejo puede apoyar y plantear las opiniones de la sociedad civil, y esta vez Meta tiene que responder”.**

Afia Asantewaa Asare-Kyei  
Miembro del Consejo



## Cómo han determinado los comentarios del público nuestras decisiones

Los comentarios del público han desempeñado un papel esencial en las 20 decisiones que el Consejo publicó en 2021. Los comentarios nos plantearon las inquietudes de los usuarios relacionadas con los casos y nos ofrecieron información lingüística, cultural y contextual que enriqueció nuestras decisiones. Durante 2021, el Consejo observó un aumento en la calidad de los comentarios del público presentados, y hubo más comentarios a los que se hizo referencia o se citó explícitamente en las decisiones publicadas posteriormente.

### Proporción de conocimientos lingüísticos

En varios casos, los comentarios del público proporcionaron contexto lingüístico crucial. En el caso *Bot de Myanmar*, los comentarios del público expresaron su acuerdo con los expertos lingüistas del Consejo de que en birmano se utilizaba la misma palabra para referirse a los estados que a la población de ese estado. Fue un contexto esencial para que el Consejo llegara a la conclusión de que, como el término ofensivo en la publicación se refería

a “China”, el estado, y no a los “chinos”, como población, el contenido debía restaurarse.

En el caso *Insultos en Sudáfrica*, los comentarios del público también ayudaron al Consejo a comprender el significado local de los insultos usados en la publicación. Contribuyeron a nuestra conclusión de que uno de los insultos era especialmente ofensivo y dañino en el contexto sudafricano, lo que, a su vez, respaldó nuestra decisión de no restaurar la publicación en Facebook.

### Oferta de mayor contexto

En nuestra decisión sobre los *confinamientos por COVID en Brasil*, los comentarios del público hicieron hincapié en la politización de las medidas para contrarrestar la propagación del COVID-19 en el país. Contribuyeron a la recomendación del Consejo de que, dada la pandemia, Meta debía priorizar la verificación de los hechos ante la posible desinformación sanitaria compartida por las autoridades públicas, teniendo en cuenta el contexto local.

### Expresión de las inquietudes de los usuarios

Los comentarios del público también facilitaron el planteamiento de inquietudes mayores relacionadas con el caso. Los comentarios presentados para la *publicación compartida de Al Jazeera*, por ejemplo, incluía alegaciones de que Meta había eliminado y relegado de forma desproporcionada contenido de usuarios palestinos y contenido en árabe, especialmente si se comparaba con el trato que daba a la violencia contra los árabes o palestinos en Israel. Estos comentarios contribuyeron a la recomendación de que urgía que Meta examinara su moderación de contenidos en árabe y hebreo se había aplicado con imparcialidad.

### Evaluación del compromiso con los derechos humanos de Meta

El Consejo recibió tres comentarios del público de Relatores especiales de la ONU, así como de otros expertos en derechos humanos de todo el mundo. En el caso *Aislamiento de Öcalan*, el Relator especial de la ONU sobre los derechos humanos



y la oposición al terrorismo indicó que las normas comunitarias de Meta y las normas comunitarias sobre las personas y organizaciones peligrosas “no son lo suficientemente coherentes con la legislación

internacional y pueden, en la práctica, mermar determinados derechos fundamentales, entre los que se incluyen la libertad de expresión, asociación, participación en actos públicos y no discriminación”.

## LECCIONES APRENDIDAS

Desde que anunciamos nuestros primeros casos en diciembre de 2020, hemos escuchado de manera regular a los usuarios y mejorado nuestro proceso de comentarios del público.

Como respuesta a los comentarios de los interesados, hemos detallado aún más los resúmenes de los casos y hemos añadido preguntas “rápidas” para destacar las áreas en las que estamos más interesados en recibir información. Para que el mayor número de personas y organizaciones puedan participar en nuestro trabajo, ampliamos el periodo original para presentar comentarios del público de 7 días a 10 días y, después, a 14 días. También incrementamos el límite de páginas para que las presentaciones puedan ser más detalladas. Aunque siempre hemos aceptado comentarios en el idioma de la publicación y en inglés, en 2022 pensamos ofrecer a los usuarios la opción de presentar comentarios en más idiomas.



# Recomendaciones



**86**

recomendaciones emitidas para Meta en 2021

Siguiendo nuestras recomendaciones, Meta se ha comprometido a:

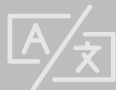


**Proporcionar más detalles**

a los usuarios a la hora de eliminar publicaciones con lenguaje que incita al odio.



Implementar nuevos sistemas de mensajería en determinadas ubicaciones para notificar a los usuarios la eliminación de contenidos como resultado de los procesos de **revisión humana** o **automatizada**.



Traducir las normas a **los idiomas de más de 400 millones de personas**.



Adoptar un nuevo **protocolo de política de crisis** para gestionar la respuesta ante situaciones de crisis.



Proporcionar **nueva información a solicitud de los poderes públicos**.



En este informe se usa un **nuevo enfoque basado en datos** para realizar un seguimiento de la implementación y garantizar que Meta honre los compromisos adquiridos al aceptar las recomendaciones.



de las 86 recomendaciones, Meta acreditó la realización de progresos o la implementación de las recomendaciones, entrando en estas categorías las recomendaciones de **transparencia**.



Las respuestas de Meta a las recomendaciones **han mejorado** con el tiempo.

# Información general

En las 20 decisiones que publicamos en 2021, realizamos 86 recomendaciones a Meta de cómo puede mejorar las políticas y los procedimientos que aplica al contenido que publican sus miles de millones de usuarios.

Aunque nuestras recomendaciones sean consultivas, Meta debe responder a ellas públicamente. En 2021, Meta disponía de 30 días desde la publicación de una decisión para responder a nuestras recomendaciones. Este plazo se ha extendido a 60 días.

Nuestras recomendaciones han instado repetidamente a Meta a que siga algunos principios fundamentales de transparencia: que las reglas estén fácilmente disponibles en los idiomas de los usuarios; que indique lo más claramente posible cómo toma y aplica las decisiones y, cuando se infrinjan, que indique al infractor exactamente lo que ha hecho de manera incorrecta.

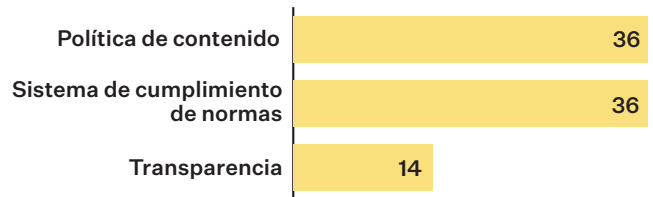
## Compromisos iniciales, creciente impacto

En 2021, Meta contrajo una serie de compromisos como respuesta a nuestras recomendaciones que, en su conjunto, tienen el potencial real de mejorar la forma en que la compañía trata a los usuarios de Facebook e Instagram en todo el mundo.

### Indicar a los usuarios las reglas y lo que han hecho mal:

- En 2021, dimos a Meta una recomendación más veces que cualquier otra y la repetimos en seis decisiones: cuando se elimine contenido de personas, **se les debe indicar qué regla específica han infringido**. Gracias a la respuesta positiva de Meta, las personas que usan Facebook en inglés reciben ahora mensajes específicos al contenido que se elimina por lenguaje que incita al odio. Meta prueba ahora aplicar este sistema de mensajería específica al contenido en árabe,

## ¿Qué tipo de recomendaciones hizo el consejo en 2021?



español y portugués, así como a las publicaciones eliminadas por *bullying* y acoso.

- En nuestras tres decisiones sobre el contenido en Instagram, instamos a Meta a que indicase al usuario que las **normas comunitarias de Facebook se aplican a Instagram**. La compañía ha dicho que está “implementando completamente” esta recomendación.
- En respuesta a nuestra recomendación de explicar y ofrecer ejemplos de términos clave en su **política de personas y organizaciones peligrosas**, Meta añadió definiciones y ejemplos de “exaltación”, “apoyo sustantivo” y “representación”. Estas clarificaciones son cruciales para que los usuarios comprendan lo que está permitido y lo que no.

### Ser transparente con los usuarios sobre cómo se modera el contenido:

- En respuesta a nuestra recomendación en la decisión sobre los *síntomas de cáncer de mama y desnudos*, Meta está implementando nuevos sistemas de mensajería en algunas ubicaciones para notificar a los usuarios la eliminación de contenido como resultado de los procesos de revisión humana o automatizada.
- Con la adopción de otra de nuestras propuestas, Meta aceptó proporcionar información sobre el contenido eliminado por infringir sus normas comunitarias **después de un informe formal de un gobierno**. Así se aumentará la transparencia sobre cómo los poderes públicos presionan a Meta y cómo responde la compañía a esa presión.

- Como respuesta a otra recomendación, Meta ha iniciado las preparaciones necesarias para notificar las instancias en las que aplica su **concesión de interés periodístico** en su informe trimestral del sistema de cumplimiento de normas comunitarias. La compañía espera implementar esto a finales de 2022.

### Tratar a los usuarios de forma equitativa, independientemente de dónde se encuentren:

- Después de otra recomendación, Meta **tradujo las normas comunitarias de Facebook al punyabí y urdu**. También se comprometió a traducir las reglas de Facebook al maratí, telugu, tamil, gujaratí y otros idiomas a principios de 2022. Una vez completada la traducción, **más de 400 millones de personas**, principalmente en Asia del Sur, dispondrán de las normas comunitarias de Facebook en su idioma nativo.
- Después de escuchar las reclamaciones de los interesados de que Meta ha eliminado de forma desproporcionada las publicaciones de palestinos, instamos a Meta a que examinase si su **moderación de contenido en hebreo y árabe era imparcial**. Aceptó hacerlo.

### Mejorar la imposición de las reglas y aplicarlas coherentemente

- En respuesta a una recomendación en nuestra decisión sobre los *síntomas de cáncer de mama y desnudos*, que trata sobre una publicación para la concienciación sobre la enfermedad que se eliminó erróneamente, Meta actualizó sus modelos de detección automática de desnudos para que tuviera en cuenta los contextos de salud.
- En nuestra decisión sobre el *aislamiento de Öcalan*, Meta primero eliminó el contenido porque decía que el usuario infringió sus reglas sobre personas peligrosas. Sin embargo, luego reveló que había encontrado **una excepción a la política que se**

**había perdido durante tres años**, por la que claramente se permitía la publicación. La excepción permitía específicamente que las personas hablasen de las condiciones de confinamiento de las personas que aparecían en su lista de personas que pertenecían a “organizaciones peligrosas”. Como resultado de nuestra recomendación en este caso, Meta ha restaurado esta instrucción. Ello significa que, en el futuro, los usuarios tendrán un mejor conocimiento de cómo modera Meta el contenido relacionado con personas peligrosas en sus plataformas.

- En respuesta a una recomendación de nuestra decisión sobre la *suspensión del expresidente Trump* y, después de consultar a más de 50 expertos globales externos a la compañía, Meta adoptó un protocolo de política de crisis para gestionar sus respuestas ante estas situaciones. Así se proporcionará una base más coherente y transparente de cómo la compañía responde en situaciones de crisis.

## Lecciones aprendidas

A pesar de los problemas con las respuestas de Meta a nuestras primeras recomendaciones, tanto Meta como el Consejo han tomado medidas para mejorar el proceso de recomendaciones durante 2021. Aunque este trabajo ya está dando sus frutos, quedan algunas preguntas clave. ¿Cómo pueden tener más peso nuestras recomendaciones? ¿Cómo podemos trabajar con Meta en esta área sin comprometer nuestra independencia? Y, fundamentalmente, ¿cómo podemos garantizar que Meta respete sus compromisos mediante acciones que el Consejo pueda medir y que las personas de todo el mundo puedan experimentar? En respuesta a estos puntos, establecimos un enfoque nuevo y más riguroso para supervisar la implementación por parte de Meta de nuestras recomendaciones, que se describen más adelante en este capítulo.

# Respuesta de Meta a nuestras recomendaciones

Meta confirmó la implementación de la mayoría de nuestras recomendaciones. 55 de 87 respuestas se encuentran en “implementación completa” o en “implementación parcial”.

En ocho de las recomendaciones, Meta respondió que estaban en “evaluación de su viabilidad”, mientras que afirmó que 14 recomendaciones ya se aplicaban. Meta dijo también que no realizaría ninguna acción más en respuesta a 10 recomendaciones. Estas cifras reflejan las respuestas de la actualización del cuarto trimestre en el Consejo, publicada en marzo de 2022. El número de respuestas de Meta (87) no coincide con el número total de recomendaciones que notificamos aquí (86) debido a las diferencias en el recuento del primer conjunto de recomendaciones del Consejo.

## Mejora del proceso de las recomendaciones

Durante 2021, vimos un progreso en las recomendaciones. El Consejo mejoró al realizarlas y Meta en su respuesta. A pesar de algunos problemas que las respuestas de Meta plantearon a nuestras primeras recomendaciones, ambas partes iteraron su enfoque. Empezamos 2022 con un mejor entendimiento del proceso de las recomendaciones y de cómo podemos mejorarlo.

En febrero de 2021, Meta respondió a las recomendaciones que hicimos en nuestras cinco primeras decisiones. Aunque se comprometió a seguir la mayor parte de nuestras propuestas, en algunos casos la compañía reestructuró o reformuló nuestras recomendaciones. En otros casos, Meta solo respondió a ciertas partes de nuestras propuestas o no pareció entender totalmente lo que intentábamos hacer.

A lo largo del primer año de trabajo, hemos cambiado también nuestros procesos, para fortalecer las recomendaciones del Consejo.

En nuestra decisión sobre la *Representación de Zwarte Piet*, creamos la nueva sección de “declaración consultiva de políticas” que enumeraba claramente



nuestras decisiones e indicaba que “el Consejo solicita que [Meta] proporcione una respuesta individual a cada una”. Desde octubre, también comenzamos con nuestros informes trimestrales de transparencia para anotar las instancias en las que Meta había tergiversado o malinterpretado nuestras recomendaciones.

También reconocemos que, por su parte, Meta adoptó medidas en 2021 para mejorar su respuesta a nuestras recomendaciones.

Con el tiempo, Meta especifica mejor sus categorías y ha sustituido “compromiso para actuar” por “implementación completa” o “implementación parcial”, por ejemplo. La compañía también pidió a BSR (Business for Social Responsibility) que explorase las opciones en la forma de interactuar con nosotros para la implementación de nuestras recomendaciones.

En general, al final del año, la compañía ofrecía respuestas más completas a nuestras recomendaciones, en comparación con el inicio del año.

Dos ejemplos de respuestas a recomendaciones realizadas a finales de 2021 ilustran este progreso. Al responder a una recomendación de nuestra decisión *Cinturón con abalorios*, Meta describió dos experimentos que había realizado en áreas relevantes



## Informe sobre la cronología de la implementación y respuesta de Meta a nuestras recomendaciones

- En los estatutos que se aplicaban en 2021, Meta tenía que responder a nuestras recomendaciones públicamente en el plazo de 30 días a partir de la publicación de una decisión del Consejo.
- En 2021, Meta respondió a nuestras recomendaciones dentro de este periodo de tiempo.

para nuestra propuesta. Con ellos, se muestra cómo el proceso de realizar recomendaciones y responder a ellas aumenta la transparencia y revela nueva información sobre los procesos internos de Meta. En una respuesta independiente a una recomendación de nuestra decisión *Bebida de ayahuasca*, Meta indicó que evaluaría la viabilidad en su foro de políticas de productos, que publica actas de sus debates. Apreciamos estas respuestas de Meta.

Meta también es ahora más transparente con el Consejo sobre la evidencia empírica de su impacto. En una reciente evaluación de viabilidad, Meta observó que si se informaba a los usuarios cuyo contenido se había eliminado debido al subtipo de infracción de lenguaje que incita al odio, comprendían mejor cómo y por qué se aplican las reglas. Meta compartió que había detectado un aumento estadístico significativo en las percepciones sobre transparencia y legitimación en varios mercados. En vista de ello, Meta experimenta ahora con el envío a los usuarios de notificaciones más específicas para las infracciones de *bullying* y acoso.

En 2021, Meta planteó en varias ocasiones problemas con el nivel de recursos necesarios para responder a nuestras recomendaciones. En su primer informe trimestral de transparencia del Consejo, publicado en julio de 2021, la compañía dijo:

*“[El] tamaño y ámbito de las recomendaciones del Consejo van más allá de las instrucciones de la política que anticipamos en primer lugar cuando establecimos el Consejo, y varias requieren inversiones de varios meses o varios años”.*

Meta expuso estos problemas en su siguiente informe trimestral de transparencia en el Consejo publicado en noviembre de 2021:

*“[E]l ritmo y volumen de las recomendaciones no nos permiten adecuar el tiempo para la evaluación e implementación inicial de las recomendaciones. [...] [C]omo promedio, nuestros equipos evalúan y responden entre 5 y 35 recomendaciones en cualquier momento. La mayoría de estas recomendaciones requieren más de una docena de personas para evaluar la viabilidad, lo que no se puede completar con facilidad en 30 días. Además, esta dificultad incluye la necesidad de incorporar las recomendaciones en las iniciativas y prioridades existentes, como las reflejadas en nuestra integridad y en los “hojas de ruta” del equipo de productos relacionados”.*

En 2021, nuestras recomendaciones presionaban a Meta a que realizase cambios ambiciosos en la experiencia de los usuarios en Facebook e Instagram, algunos de los cuales podían requerir realizar cambios técnicos significativos. Sabemos que primero se necesita tiempo para evaluar si los cambios son viables y así poder proporcionarnos una respuesta completa. En respuesta a estas inquietudes, aceptamos ampliar el plazo para que Meta responda a nuestras recomendaciones de 30 a 60 días.

Sin contar los casos de la *suspensión del expresidente Trump* y del *aislamiento de Öcalan*, en que Meta solicitaba específicamente instrucciones de política detalladas, las decisiones publicadas en 2021 incluían, de promedio, alrededor de tres recomendaciones cada una. Así que nos hemos propuesto garantizar que las recomendaciones del Consejo sean implementables y

que se implementen, y hemos hecho partícipe a Meta en la dotación de los recursos necesarios para actuar.

## Mejora de las respuestas de Meta a nuestras recomendaciones

Este informe contiene nuevos análisis del Consejo de las respuestas de Meta a las recomendaciones que hicimos durante 2021. La medición de la “respuesta” evalúa si la respuesta de Meta a la recomendación del Consejo fue completa. Estas estadísticas reflejan las actualizaciones trimestrales de Meta en el Consejo durante el tercer trimestre de 2021.

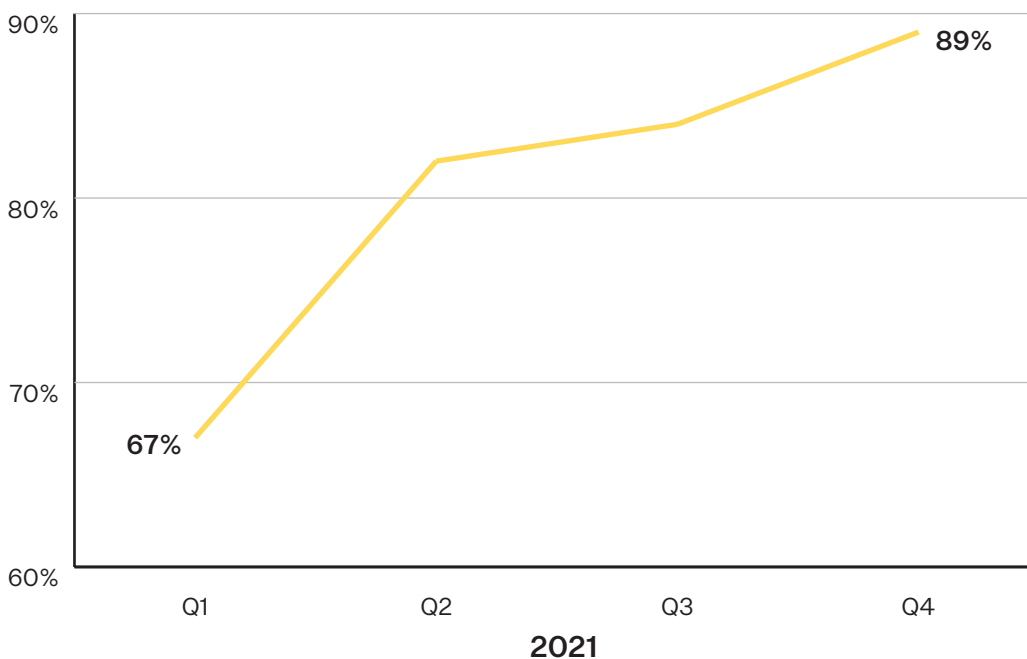
Al evaluar cada respuesta, formulamos tres preguntas: (1) ¿Aborda todos los aspectos de nuestra recomendación? (2) ¿Ofrece un compromiso para

actuar? (3) ¿Proporciona una cronología? Cuando Meta cumplía uno de estos criterios consideramos que había proporcionado una respuesta “bastante completa” y cuando cumplía al menos dos, consideramos que había proporcionado una respuesta “completa”.

Cuando examinamos las respuestas de Meta a las recomendaciones publicadas en distintos trimestres, podemos ver que la proporción de respuestas consideradas “completas” o “bastante completas” aumentó en cada periodo, subiendo del 67% en el primer trimestre al 89% en el cuarto trimestre de 2021.

Aunque aún hay margen de mejora, parece que los cambios en el proceso de recomendaciones de 2021 ha dado como resultado respuestas más completas de Meta.

### ¿Con qué frecuencia ofreció Meta una respuesta “completa” o “bastante completa” a nuestras recomendaciones?



# De compromisos a la acción: resultados para los usuarios



“Es importante garantizar que los compromisos adquiridos por Meta, algunos de los cuales tardarán en cumplirse, no se abandonan y olvidan”.

Director del Consejo asesor de contenido, Thomas Hughes



Aunque Meta se comprometió a implementar la mayoría de nuestras recomendaciones de 2021, nuestra siguiente tarea es garantizar que mantiene sus promesas. La prioridad principal del Consejo es garantizar que Meta convierte sus compromisos en acciones que mejoren la forma en que la compañía trata a las personas de todo el mundo.

Al crecer el número de recomendaciones durante 2021, ampliamos nuestro trabajo a la implementación. Este informe establece un enfoque riguroso basado en datos, que nos permite hacer el seguimiento de cómo Meta realiza la implementación de las recomendaciones y poder evaluar con claridad su impacto en los usuarios.

## Desarrollo de nuestra capacidad de pedir responsabilidades a Meta

Desde mediados de 2021, empezamos ampliar nuestra capacidad de supervisar la implementación por parte de Meta de las recomendaciones. En julio, creamos el equipo de implementación y supervisión de casos y lo dotamos de personal para apoyar al Consejo. Este equipo, que se ha ampliado desde entonces, supervisa y mide las respuestas y acciones de Meta para comprender el impacto de nuestras recomendaciones en los usuarios de Facebook e Instagram.

A mediados de 2021, también establecimos un grupo de trabajo de implementación compuesto por miembros del Consejo y personal experimentado de Meta. Este grupo se reunió varias veces para discutir y mejorar el proceso de recomendaciones, y Meta respondió a nuestras preguntas sobre sus procesos para implementar nuestras propuestas. Ese mismo año, también establecimos un comité de implementación, que actualmente se compone de cinco miembros del Consejo, para reunirse con nuestros comités de selección de casos y de miembros. Esta acción representó una clara elección de colocar la implementación al mismo nivel que las funciones más importantes de nuestra organización. El comité de implementación se ha esforzado en refinar las recomendaciones del Consejo y en garantizar que se centran en impactos específicos y medibles.

## Un nuevo enfoque basado en datos de la implementación

Este informe representa otro hito en nuestra andadura para pedir responsabilidades a Meta. En el anexo de este informe proporcionamos, por primera vez, nuestra evaluación de la implementación por parte de Meta de cada una de las 86 recomendaciones realizadas en 2021. Estas estadísticas reflejan las actualizaciones de Meta en el tercer trimestre de 2021.

Para medir el progreso de Meta en la implementación, miramos si se han cumplido ciertos criterios de una recomendación determinada. Estos criterios varían en función de la recomendación. Por ejemplo, si propusimos que Meta añadiera más detalles a sus normas comunitarias públicas o emitiera un informe sobre la diligencia debida sobre los derechos humanos en un determinado tema, su publicación externa demuestra su implementación. Para otras recomendaciones, Meta debe suministrar datos que no están disponibles públicamente para demostrar su implementación. Ello significa que debe proporcionar los datos de notificación a los usuarios, para probar que se indica a los usuarios qué regla



específica infringieron. En el futuro, mediremos la implementación que hace Meta de nuestras recomendaciones según cuatro categorías y la actualización trimestral de nuestras evaluaciones:

- *“Implementación demostrada”*: Meta ha proporcionado suficientes datos para poder decir que esta recomendación se ha completado.
- *“Progreso realizado”*: Meta se ha comprometido a implementar esta recomendación, pero aún no ha completado todas las acciones necesarias.
- *“Meta ha informado de la implementación o ha descrito el trabajo que ya realiza, pero no ha publicado información para acreditar la implementación”*: Meta dice que ha implementado su recomendación, pero no ha proporcionado suficiente evidencia para que podamos verificarlo.
- *“Recomendación omitida, rechazada o reformulada”*: Meta no realizará ninguna acción más a cuenta de nuestra propuesta.

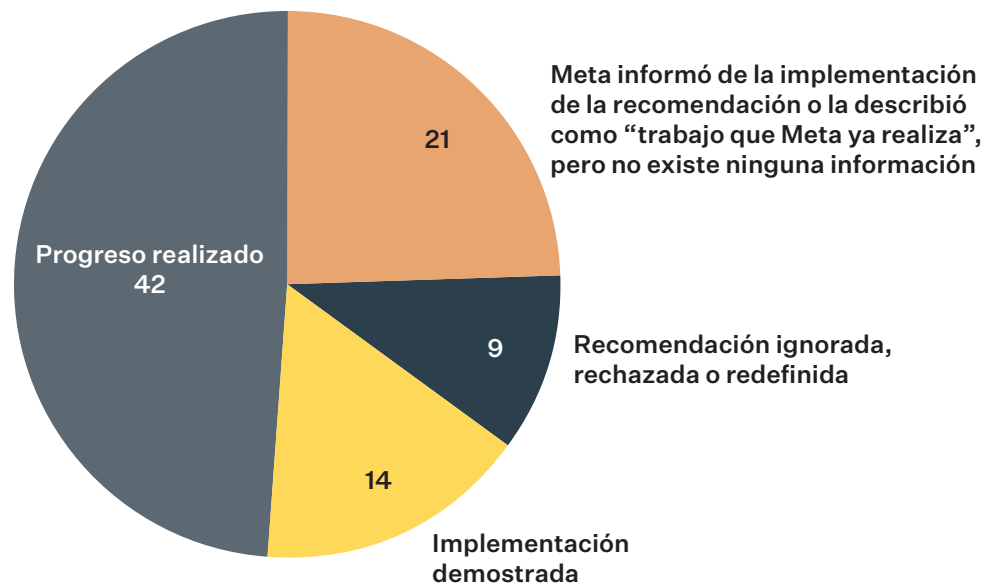
Este nuevo enfoque basado en datos significa que, a veces, la evaluación de si Meta ha implementado una recomendación puede diferir de los informes de la compañía. Sin embargo, creemos que esta clase de validación independiente es crucial

para pedir responsabilidades a Meta y garantizar que los usuarios experimentan el impacto de nuestras recomendaciones. En el siguiente gráfico se muestra un desglose de nuestra evaluación de la implementación por parte de Meta de las recomendaciones:

Ello muestra que en 2021 Meta demostró la implementación de 14 recomendaciones. Dado que el Consejo solo ha hecho recomendaciones durante un periodo relativamente breve de tiempo y debido al grado de ambición de las propuestas del primer año, Meta tardará en completar nuestras recomendaciones y demostrar que lo ha hecho. Del total de recomendaciones, 42 se encuentran en la categoría de “progreso realizado”, que incluye muchas de las recomendaciones de alto impacto propuestas en la introducción de esa sección. Esperamos que muchas recomendaciones de esta categoría pasen a “implementación demostrada” durante 2022.

En el caso de 21 recomendaciones, aunque Meta notificó que había implementado nuestra propuesta o la describió como un trabajo que ya realizaba, no publicó suficiente información para que pudiésemos validar la implementación. Además, animamos a Meta a que comparta más datos y proporcione más

## Implementación de Meta de nuestras recomendaciones



evidencias para respaldar las declaraciones de que ha implementado nuestras recomendaciones. También alentamos a Meta a que conceda mayor acceso a los datos de la compañía para permitir validar el progreso. Nueve recomendaciones se encontraban en la categoría de “recomendación omitida, rechazada o reformulada”, lo que significa que Meta no realizará ninguna acción más a cuenta de estas propuestas.

Cuando analizamos los diferentes tipos de recomendaciones evaluadas como “implementación demostrada” o “progreso realizado”, encontramos que el 79% de las recomendaciones de transparencia y el 67% de las recomendaciones de política sobre contenidos se encontraban en esas categorías, en comparación con el 58% de las recomendaciones del sistema de cumplimiento de normas. La menor proporción de recomendaciones del sistema de cumplimiento de normas puede reflejar que estas propuestas requieren con frecuencia que Meta realice cambios mayores en Facebook e Instagram.

## Obtención de resultados para los usuarios de Facebook e Instagram

Aunque la validación de la implementación que Meta hace de nuestras recomendaciones es una tarea enormemente compleja, hemos elegido priorizar e invertir en esta área. El motivo de estas inversiones es sencillo: para mejorar la forma en que Meta trata a los usuarios de Facebook e Instagram a largo plazo y garantizar que respeta sus responsabilidades para con los derechos humanos, debemos ayudar a la compañía a convertir sus compromisos en acciones.



# Petición de responsabilidades a Meta sobre la verificación cruzada

El 13 de septiembre de 2021, The Wall Street Journal informó que el programa de verificación cruzada de Meta era mucho más complejo en escala y ámbito de lo que la compañía había informado previamente. Uno de los primeros artículos de una serie que se basaba en los documentos de la compañía obtenidos de un denunciante señalaba:

*Mark Zuckerberg ha dicho públicamente que [Meta] permite que sus más de tres mil millones de usuarios hablen en el mismo nivel que las élites políticas, culturales y periodísticas y que sus normas de comportamiento se aplican a todos, sin importar su estatus o fama. [Pero] en privado, la compañía ha construido un sistema que exime a los usuarios de alto perfil de algunas o de todas estas reglas.<sup>7</sup>*

El artículo afirmaba que la verificación cruzada se había ampliado desde un programa que “inicialmente se preveía como una medida de control de calidad de las acciones realizadas contra las cuentas de perfiles de alto nivel, incluidas las celebridades, los políticos y los periodistas” a un programa que protegía a “millones de usuarios VIP” de los procesos del sistema de cumplimiento de normas habitual de la compañía. Según los documentos del denunciante, algunos publicadores con un perfil de alto nivel se incluyeron en la lista de aprobados. El artículo informaba de que parte de este proceso permitía que empleados individuales de Meta añadieran determinadas entidades (por ejemplo, perfiles o páginas) a listas que supuestamente permitían que “material que infringía reglas” permaneciera de manera indefinida, “pendiente de revisiones de empleados de Facebook que con frecuencia nunca llegaba”.<sup>8</sup>

El 21 de septiembre de 2021, publicamos nuestra respuesta a las manifestaciones:

*En el Consejo asesor de contenido, hemos formulado preguntas sobre la verificación cruzada durante un periodo de tiempo. En nuestra decisión relativa a las cuentas del expresidente de los Estados Unidos Donald Trump, advertimos que la falta de información pública clara sobre la verificación cruzada y la “excepción de interés periodístico” de [Meta] podía contribuir a la percepción de que [la compañía] está excesivamente influenciada por consideraciones políticas y comerciales.*

Entre las recomendaciones que enviamos a Meta como parte de esa decisión, una le instaba a “informar sobre la proporción de errores relativos y la coherencia temática de las determinaciones realizadas mediante el proceso de verificación cruzada en comparación con los procedimientos ordinarios del sistema de cumplimiento de normas”. En su respuesta, Meta rechazó implementar esta propuesta sobre la base de que “el seguimiento de esta información no es viable”. Meta había hecho copia de seguridad de esta reclamación con un vínculo a una publicación de blog de 2018 que indicaba: “Eliminamos contenidos de Facebook, independientemente de quién lo publique, cuando infringe nuestras normas”. Sin embargo, The Wall Street Journal, informó que una revisión interna de 2019 realizada por Meta estimó que la publicación del blog era “engañoso”.<sup>9</sup>

El Consejo concluyó, “A raíz de los desarrollos recientes, estamos investigando en qué grado [Meta] ha sido clara en sus respuestas en relación con la verificación cruzada, incluido el proceso de la lista blanca. El Consejo se ha puesto en contacto con [Meta] para (...) esclarecer más la información que se compartió anteriormente con nosotros. Esperamos recibir una sesión informativa o briefing en los próximos días.”

<sup>7</sup> “Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That’s Exempt,” Jeff Horwitz, *The Wall Street Journal*, septiembre. 13, 2021

<sup>8</sup> Ibid.

<sup>9</sup> Ibid.

En esa sesión informativa, el equipo de Meta ha admitido que “no deberíamos haber dicho que la verificación cruzada solo se aplicaba a ‘un número pequeño de decisiones’ en la respuesta a nuestra recomendación en el caso relacionado con el expresidente Trump. El 28 de septiembre del 2021, una semana después de nuestra publicación del blog del 21 de septiembre, Meta respondió a las preguntas del Consejo enviando su propia solicitud para obtener una opinión consultiva sobre políticas. Meta pidió que incluyéramos recomendaciones específicas sobre cómo mejorar la verificación cruzada. Al anunciar esta recomendación, Meta añadió:

*Recientemente, el Consejo ha mostrado interés en profundizar en nuestro sistema de verificación cruzada. Esta recomendación va más allá de la sesión informativa que ya se había hecho. Estamos pidiendo de manera proactiva al Consejo que nos dé sus opiniones mediante un proceso formal y transparente... Concretamente, le pediremos que nos guíe en los criterios que usamos para determinar qué se prioriza en una revisión secundaria con la verificación cruzada y la forma de gestionar el programa.<sup>10</sup>*

A principios de octubre, antes de publicar el primer informe trimestral de transparencia del Consejo, anunciamos que nos reuniríamos con Frances Haugen, una informante extrabajadora de Meta, para tratar los importantes problemas que planteó en sus revelaciones sobre el proceso y las políticas de moderación de contenido de Meta.

*En las últimas semanas, se ha dado a conocer nueva información sobre el enfoque de [Meta] en la moderación de contenido a raíz de las acciones de la extrabajadora de [Meta], Frances Haugen. A causa de las graves afirmaciones que Haugen ha hecho sobre [Meta], la hemos invitado para que hable con el Consejo en las próximas semanas y ha aceptado la propuesta. Los miembros del Consejo agradecen*

*la oportunidad de hablar con Haugen sobre sus experiencias y recopilar información que puede motivar una mayor transparencia y responsabilidad por parte de [Meta], gracias a nuestras decisiones de casos y recomendaciones.*

Nuestro blog concluyó diciendo:

*Las decisiones que toman empresas como [Meta] tienen consecuencias en el mundo real, ya que afectan a la libertad de expresión y a los derechos humanos de miles de millones de personas en el mundo. En este contexto, la transparencia de las normas es esencial.*

*Como Consejo, seguiremos haciendo preguntas difíciles a Facebook y presionaremos a la empresa para que se comprometa a ofrecer una mayor transparencia y que así sea más responsable y justa. A la larga, solo esta opción permitirá que los usuarios sepan que se les está tratando de manera justa.<sup>11</sup>*

En el primer informe trimestral de transparencia que se publicó a finales de ese mes manifestamos que, en nuestra opinión, el equipo interno de Meta que tenía la tarea de proporcionar información no ha sido “plenamente sincero” sobre la verificación cruzada. “En algunas ocasiones”, declaramos, “[ese equipo] no proporcionó al Consejo información relevante (...) [mientras que] en otras instancias, la información que proporcionó estaba incompleta.” A raíz de ello, manifestamos que la futura relación laboral con Meta depende de nuestra capacidad de “confiar en que la información que nos proporciona Meta del tema en cuestión es precisa, exhaustiva y completa”. Para cumplir con ese objetivo, el 21 de octubre del 2021 anunciamos que habíamos aceptado la solicitud de Meta de crear una opinión consultiva sobre políticas que revisaría el programa de verificación cruzada y que enviaríamos recomendaciones sobre cómo se podría mejorar.

10 “Requesting Oversight Board Guidance on Our Cross-Check System,” publicación de blog de Meta de Nick Clegg, VP, Global Affairs, 28 de septiembre del 2021 (actualizado el 21 de octubre del 2021)

11 “Oversight Board to meet with Frances Haugen” Publicación en el blog del Consejo asesor de contenido en octubre de 2021

# Participación y alcance

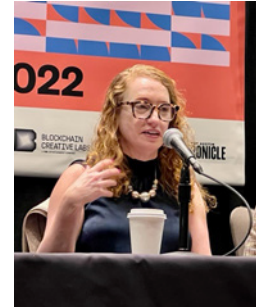
En nuestro primer año, hemos trabajado con personas y organizaciones de todo el mundo, hemos proporcionado información sobre el trabajo que desempeña el Consejo, hemos animado a los usuarios a que envíen su opinión sobre nuestras decisiones, y hemos debatido por qué un enfoque ejemplar y global en la moderación de contenidos es importante para los usuarios.



Julie Owono



Nighat Dad



Suzanne Nossel

## Animamos a los usuarios a que compartan su opinión

Gracias a la participación externa, muchas personas y organizaciones se han animado a colaborar con nosotros y a enviar comentarios del público sobre los casos en curso. Hemos celebrado 19 sesiones en horarios de oficina para explicar cómo se pueden presentar los comentarios del público, y han participado en ellas más de 500 partes interesadas. Para fomentar las contribuciones que llegan de todo el mundo, estas sesiones se produjeron en diversas zonas horarias y ofrecían la posibilidad de realizarlas en lenguas como el árabe, el hebreo, el portugués o el español. También hemos llevado a cabo 12 eventos y mesas redondas en dos semanas tras anunciar el caso para promover que se hagan comentarios del público en el caso sobre la *suspensión del expresidente Trump*. Han asistido más de mil personas a estas sesiones.

## Impulso del debate sobre moderación de contenido y derechos humanos

Nuestro enfoque sobre cómo aplicar los derechos humanos a la moderación de contenidos no se ha formado solo a partir del debate entre los miembros del Consejo. También se ha nutrido de discusiones públicas con representantes de la academia, el periodismo, la sociedad civil y de las instituciones internacionales.

En junio celebramos eventos en RightsCon, incluida una sesión para alinear la moderación de contenido con los derechos humanos. En septiembre, durante la Asamblea General de las Naciones Unidas, la miembro del Consejo Julie Owono se unió a dos paneles de la ONU sobre cómo contrarrestar la desinformación y el lenguaje que incita al odio, y cómo promover la transparencia. En diciembre, la miembro del Consejo y antigua relatora especial de la ONU habló en un evento sobre qué lecciones había aprendido el Consejo sobre la protección de los derechos humanos y la gobernanza de las plataformas digitales. Al hablar con Peggy Hicks, directora del Consejo de Derechos Humanos de las Naciones Unidas, Maina Kiai dijo, “En el Consejo asesor de contenidos, sabemos que abordamos solo una parte del problema. Expertos de otras muchas áreas tienen que actuar también”. Asimismo, en diciembre organizamos un evento en el Foro para la Gobernanza de Internet. A este evento asistieron el director del Consejo Thomas Hughes, la miembro del Consejo Afia Asantewaa Asare-Kyei y el fideicomisario del Consejo Cherine Chalaby y en él se debatieron temas sobre los derechos digitales y la autorregulación.

## Globalización del debate sobre moderación de contenidos

Las redes sociales influyen en acontecimientos y en miles de millones de personas en el mundo. Sin embargo, en muchas ocasiones, el debate sobre la moderación de contenidos solo tiene en cuenta a

usuarios en Europa y América del Norte. Para que todo el mundo pueda formar parte de este debate, hemos priorizado los eventos que analizaban los retos de moderación de contenidos a los que se enfrentan en el Sur Global.

En mayo, los miembros del Consejo Catalina Botero-Marino y Jamal Green participaron en un evento denominado Diálogo Inter-Americano que trató sobre qué significado tienen las decisiones del Consejo en el Sur Global y, en particular, en América Latina. En agosto, el miembro del Consejo Endy Bayuni tomó parte en un evento con FORUM Asia sobre la ciberseguridad centrada en la humanidad. En septiembre, la miembro del Consejo Julie Owono abordó las lecciones sobre la moderación de contenido de África Occidental y Central. En ese evento asistió la Comisión de Protección de Datos Personales de Senegal.

Por último, en noviembre los miembros del Consejo Nighat Dad, Ronaldo Lemos y Maina Kiai hablaron en

“

**Lo que se propone en el Norte Global puede causar daños a comunidades del Sur Global”.**

**Nighat Dad**  
*Miembro del Consejo*



el Foro por la paz de París (Paris Peace Forum). Nighat Dad señaló que “Alguna de las regulaciones que vemos en el Sur Global son idénticas a las del Norte Global y no tienen en cuenta el contexto local ni las realidades de nuestra región. Lo que se propone en el Norte Global puede causar perjuicios a comunidades del Sur Global.”



# Próximos pasos: 2022 en adelante

En 2021, el Consejo asesor de contenido dio a conocer 20 decisiones e hizo 86 recomendaciones a Meta. En su conjunto muestran nuestro compromiso de pedir responsabilidades a la empresa y conseguir una mayor transparencia por su parte. A medida que pase 2022, además de añadir nuevos miembros en el Consejo, tenemos el objetivo de aumentar nuestro impacto y mejorar cómo trata Meta a las personas y a las comunidades del mundo.

## Ampliación de nuestro alcance

Estamos dialogando con Meta para expandir el alcance del Consejo, incluida la revisión de las apelaciones de los usuarios contra las decisiones de la empresa en ámbitos como los grupos y las cuentas, para que no se tengan solo en cuenta las publicaciones individuales. Gracias a esto, esperamos poder informar de los progresos que hayamos conseguido durante

el próximo año. También estamos revisando nuestro rol en la evaluación de los planes de moderación de contenidos en el “Metaverso” emergente. En 2021, la empresa invirtió diez mil millones en el Metaverso, cosa que indica que será un ámbito en el que centrarse en los años venideros.

## Efectos a nivel global

El Consejo debe tener un alcance global, ya que el impacto de Meta es palpable en todo el mundo. En 2021, más de dos tercios de las apelaciones de los usuarios provinieron de Europa, Estados Unidos y Canadá: un 49% de Estados Unidos y Canadá, y un 20% de Europa. En cambio, solo el 14% fueron de América Latina o el Caribe; el 9% de Asia-Pacífico y Oceanía; el 4% de Oriente Medio y África del Norte; el 2% de Asia Central y del Sur; y un 2% de África subsahariana. Esta distribución refleja un desajuste en el que estamos trabajando.

En 2022 hemos aumentado nuestro alcance y la participación de las partes interesadas para trabajar con responsables políticos y líderes de la sociedad civil en Asia, América Latina, el Oriente Medio y África. Estamos construyendo una red global de consultores regionales para animar a la gente a presentar apelaciones y comentarios del público en sus respectivas regiones. También tenemos la intención de aumentar el número de lenguas en las que aceptamos comentarios del público, aparte del inglés y la lengua del contenido en cuestión.

## Publicación de opiniones consultivas sobre políticas

Además de revisar casos individuales para eliminar o restaurar contenidos, el Consejo también puede aceptar solicitudes de opiniones consultivas sobre políticas de Meta. Estas nos permiten revisar las políticas de la empresa en una área en concreto y hacer recomendaciones de cómo pueden mejorarse. En 2021, aceptamos dos solicitudes de este tipo de

Meta sobre las opciones para compartir información residencial privada y sobre el sistema de verificación cruzada de la empresa. Nuestra opinión sobre las opciones para compartir información residencial privada se publicó en febrero de 2022 y la opinión sobre la verificación cruzada se publicará a finales de este año.

## Garantizar un mayor acceso a los datos

En 2022, como parte de nuestro trabajo para garantizar que Meta respete sus compromisos, el Consejo está intentando tener acceso a los estudios y datos de Facebook. Este acceso mejoraría los casos que selecciona el Consejo, así como las

recomendaciones que hacemos y el impacto que tenemos. También tenemos en mente contratar científicos de datos como parte nuestros esfuerzos para entender y ampliar nuestro impacto.

## Compartir lo que hemos aprendido hasta ahora

En nuestro primer año, desarrollamos, mejoramos e implementamos un nuevo enfoque para la moderación de contenido gracias a la aplicación de un marco de referencia basado en las normas internacionales sobre derechos humanos. De esta manera, pudimos crear un conjunto de prácticas recomendadas y adquirimos una gran cantidad de experiencia.

En 2022, esperamos poder compartir lo que hemos aprendido en nuestro esfuerzo conjunto para mejorar las

redes sociales. También seguiremos ejerciendo nuestro rol como fuente de autorregulación en la evolución general del ámbito de la moderación de contenidos y la regulación. Compartiremos nuestra capacidad para encontrar soluciones de autorregulación en conversaciones de políticas tecnológicas, además de contribuir en diálogo sobre innovación en la moderación de contenidos con socios potenciales del sector de la tecnología.

## Creación de una institución para garantizar el éxito a largo plazo

En último lugar, los copresidentes y los miembros del Consejo, los fideicomisarios y la administración continuarán construyendo una institución de alto rendimiento y sostenible que se guíe mediante un plan estratégico a más de un año. Esta visión incluirá procesos autónomos de identificación, selección e incorporación de

miembros del Consejo, así como de su rendimiento y conducta. También continuaremos garantizando la gestión eficiente y responsable de los recursos del Fondo, lo que incluye la presentación de un modelo de financiación a largo plazo.

## Conclusión

La información de nuestro primer informe anual muestra que el Consejo asesor de contenido ya está teniendo un impacto tangible en el trabajo de Meta y en la manera en que asiste a sus usuarios. Todas las decisiones que hemos tomado, la experiencia que hemos adquirido y las conclusiones a las que hemos llegado son solo los primeros pasos de lo que será un viaje mucho más largo.

El término “moderación de contenidos”, aunque puede parecer anodino, esconde una gran verdad: el contenido que se publica y comparte en las redes sociales puede tener un gran impacto (positivo o negativo) en la vida de las personas tanto en la red como fuera de ella. En general, aunque las redes sociales han alcanzado su potencial pronosticado de reunir a miles de millones de personas, también ha creado maneras nuevas de causar daño a los demás.

El impacto del Consejo no solo se notará cada vez más en los resultados de decisiones de contenido concretas, sino que los usuarios se beneficiarán de una aplicación

clara, justa y transparente de las propias políticas de Meta. Esperamos que el trabajo del Consejo brinde un marco de referencia fidedigno para otras empresas de redes sociales, editores de plataformas, reguladores y responsables políticos del mundo que se enfrentan a las tensiones y los retos intrínsecos de la moderación de contenidos.

En el Consejo asesor de contenido presentamos este primer informe anual como una exposición precisa de ese trabajo. Estamos orgullosos de haber llegado tan lejos, pero sabemos que aún tenemos una cantidad sustancial de trabajo por delante. Este documento refleja, principalmente, nuestro compromiso colectivo para fomentar los valores de transparencia y responsabilidad que harán de las redes sociales un lugar mejor para que los usuarios pasen el tiempo. Seguiremos pidiendo responsabilidades a Meta con el objetivo de mejorar cómo trata a la gente y a las comunidades en todo el mundo.



# Anexo

## Cómo respondió Meta a nuestras recomendaciones y cómo las implementó<sup>12</sup>

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Armenios en Azerbaiyán</b>				
1	Garantizar que siempre se notifique a los usuarios de cualquier sistema de cumplimiento de normas de las Normas comunitarias que estén incumpliendo, incluida la regla específica que Facebook aplique. Al hacer esto, Facebook podrá alentar expresiones que cumplan las normas comunitarias, en lugar de adoptar una posición adversa respecto a los usuarios. En este caso, se informó al usuario de que la publicación infringía la norma comunitaria sobre el lenguaje que incita al odio. Sin embargo, no se le notificó que esta también infringía las normas comunitarias porque incluía un insulto que hacía referencia la nacionalidad de origen. Facebook cumplió el principio de legalidad en esta instancia, pero su falta de transparencia hizo que su decisión fuera susceptible a la interpretación errónea de que se había eliminado la publicación porque el usuario habló de un tema controvertido o expresó un punto de vista con el que la empresa discrepaba.	E	●	■
<b>Síntomas del cáncer de mama y desnudos</b>				
1	Mejorar la detección automática de imágenes con texto superpuesto para garantizar que las publicaciones que generen conciencia sobre el cáncer de mama no se marquen erróneamente para revisión.	E	●	■
2	Garantizar que siempre se notifique a los usuarios por qué se han aplicado las políticas sobre contenidos en su contra, proporcionando para ello la regla específica de las normas comunitarias según la cual Facebook basó su decisión.	E	✘	■
3	Informar a los usuarios cuando se aplica la automatización del sistema de cumplimiento de normas en sus contenidos (se incluyen descripciones accesibles de qué significa este proceso).	E	●	■
4	Garantizar que los usuarios puedan apelar decisiones que tomen los sistemas automatizados para que se haga una revisión humana cuando se les informe que su contenido ha infringido las normas comunitarias de Facebook sobre desnudos y actividad sexual de adultos. Cuando Facebook tenga el objetivo de prevenir la explotación sexual infantil o la difusión no consentida de imágenes íntimas, debe hacer cumplir las normas comunitarias sobre explotación sexual de adultos e infantil, y sobre abuso y desnudos en lugar de imponer en exceso las políticas de desnudez adulta. De todas maneras, en este tipo de casos se deben permitir las apelaciones para que las eliminaciones incorrectas de desnudez adulta consensuada se pudieran revertir.	E	✘	■
5	Implementar un proceso de revisión interno para analizar de forma continuada una muestra representativa estadística de las decisiones automatizadas de eliminación de contenido que se han revocado y aprender de los errores de implementación.	E	✘	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

### IMPLEMENTATION

■ Implementación demostrada con información publicada

■ Progreso realizado

■ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Breast cancer symptoms and nudity</b>				
6	Ampliar la transparencia para exponer datos del número de decisiones automatizadas de eliminación por las normas comunitarias y desvelar qué proporción de dichas decisiones se revocaron una vez hecha la revisión humana.	T	●	■
7	Revisión de la explicación “corta” de las normas comunitarias de Instagram para aclarar que la prohibición de la desnudez adulta no es absoluta.	CP	✘	■
8	Revisión de la explicación “larga” de las normas comunitarias de Instagram para aclarar que se pueden mostrar los pezones femeninos para generar concienciación sobre el cáncer de mama.	CP	●	■
9	Aclaración de que las normas comunitarias de Instagram se interpretan de la misma manera que las de Facebook y, cuando hay inconsistencias, prevalece la más reciente.	CP	●	■
<b>Cita nazi</b>				
1	Cerciorarse de informar siempre a los usuarios de los motivos por los que se aplica cualquier medida en su contra para garantizar el cumplimiento de las normas comunitarias e indicar la regla específica que intenta hacer cumplir (por ejemplo, cuando se apoya a una organización que promueve el odio).	E	●	■
2	Explicar los términos clave de la política sobre personas y organizaciones peligrosas, así como brindar ejemplos de su aplicación, incluido el significado de “exaltación”, “apoyo” y “representación”. Deben coincidir con las definiciones incluidas en las normas de implementación internas de Facebook. Las normas comunitarias deben ofrecer pautas más claras a los usuarios sobre cómo reflejar de forma evidente su intención al debatir sobre personas u organizaciones consideradas peligrosas.	CP	●	■
3	Proporcionar una lista pública de las personas y organizaciones consideradas peligrosas en virtud de las normas comunitarias sobre personas y organizaciones peligrosas, o bien, como mínimo, ofrecer una lista de ejemplos. Así los usuarios podrán entender mejor la política y modificar su conducta en consecuencia.	CP	●	■
<b>Supuesta cura del COVID</b>				
1	El Consejo recomienda que Facebook establezca una norma comunitaria clara y accesible sobre información errónea relacionada con la salud. De este modo, consolidaría y aclararía las reglas actuales en un único lugar (ello incluye definir términos clave, como “desinformación”). La elaboración de esta regla debe incluir “casos hipotéticos detallados que ilustren los matices de interpretación y aplicación de [estas] reglas”, a fin de ofrecer mayor claridad a los usuarios (consulta el informe A/HRC/38/35, párrafo 46 (2018)). Facebook, de forma conjunta con partes interesadas relevantes, debe realizar una evaluación sobre la repercusión de sus decisiones en los derechos humanos como parte del proceso de actualización de reglas (UNGP, principios 18 y 19).	CP	●	■
2	Para garantizar que esas medidas contra la desinformación sobre la salud representen los métodos menos intrusivos para proteger la salud pública, el Consejo recomienda a Facebook que haga lo siguiente: Aclarar los daños específicos que quiere prevenir y ser transparente sobre cómo gestionará el daño potencial de un contenido en particular; hacer una evaluación sobre la gama de herramientas existente para hacer frente a la desinformación; considerar el potencial para desarrollar más herramientas menos intrusivas que la eliminación de contenido; publicar un abanico de opciones del sistema de cumplimiento de normas referente a las normas comunitarias y ordenarlas de más a menos intrusivas según cómo vulneren el derecho a la libertad de expresión; explicar qué factores, incluidos los criterios basados en evidencias, usará la plataforma para seleccionar la opción menos intrusiva cuando haga cumplir las normas comunitarias para proteger la salud pública; aclarar en las normas comunitarias qué opción del sistema de cumplimiento de normas aplicará a cada norma.	E	●	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

- Implementación demostrada con información publicada
- Progreso realizado

- Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación
- Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Supuesta cura del COVID</b>				
3	En casos en los que los usuarios publican información sobre tratamientos del COVID-19 que contradicen los consejos específicos de las autoridades sanitarias y cuando se ha detectado que puede haber perjuicio no inminente, el Consejo recomienda encarecidamente que Facebook adopte un conjunto de medidas menos intrusivas. Entre ellas, se encuentra el etiquetado, que alerta a los usuarios sobre la naturaleza conflictiva del contenido de una publicación y proporciona enlaces a las opiniones de la Organización Mundial de la Salud y las autoridades sanitarias nacionales. En determinadas situaciones, tal vez sea necesario implementar más obstáculos en una publicación; por ejemplo, evitando que las personas interactúen con ella o la compartan, a fin de reducir la amplificación orgánica o generada por algoritmos. También se puede bajar de categoría el contenido para evitar que otros usuarios lo vean en la sección de noticias. Los usuarios deben conocer con claridad todas las medidas de aplicación de normas, como el etiquetado y otros métodos para poner obstáculos y deben saber que estas están sujetas a apelación.	E	✘	
4	Debe publicar un informe de transparencia sobre cómo se han aplicado las normas comunitarias durante la crisis sanitaria mundial que ha ocasionado la pandemia de COVID-19. El informe debe incluir lo siguiente: Datos en términos totales y porcentuales sobre la cantidad de eliminaciones, así como información sobre otras medidas de cumplimiento de normas, sobre las Normas comunitarias específicas aplicadas, incluso sobre la proporción de eliminaciones que se realizaron completamente por medio del proceso de automatización; un desglose del tipo de contenido sobre el que se aplicaron las normas (como son publicaciones, cuentas y grupos individuales); un desglose de la fuente de detección (como es automatización, denuncia de un usuario, socios de confianza, autoridades policiales); un desglose por región e idioma; métricas sobre la eficacia de medidas menos invasivas (por ejemplo, el impacto del etiquetado o el cambio a una categoría inferior); datos sobre la disponibilidad de apelaciones durante la situación de emergencia, lo que incluye el número total de casos en que se retiró una apelación en su totalidad, y el porcentaje de apelaciones automatizadas; conclusiones obtenidas y lecciones aprendidas, como información sobre los cambios que Facebook realiza para asegurar un mayor cumplimiento con sus responsabilidades en cuanto a los derechos humanos de cara al futuro.	T	✘	
<b>Protestas en India contra Francia</b>				
1	Para garantizar que los usuarios tienen claro qué contenido está permitido, el Consejo recomienda que Facebook dé a los usuarios información adicional sobre el alcance y el sistema de cumplimiento de esta norma comunitaria. Los criterios del sistema de cumplimiento de normas deben ser públicos y ser coherentes con los Estándares de implementación internos de Facebook. Más concretamente, los criterios de Facebook deben hacer referencia a la intención, a la identidad del usuario y de la audiencia y al contexto.	CP	●	
<b>Representación de Zwarte Piet</b>				
1	Facebook debe enlazar la norma comunitaria sobre el lenguaje que incita al odio que prohíbe el estilo “blackface” (la cara pintada de negro) con el razonamiento de la empresa, incluidos los daños que quiere prevenir.	CP	●	
2	En la misma línea que la recomendación del Consejo en el caso 2020-003-FB-UA, Facebook debe “garantizar que siempre se notifique a los usuarios los motivos por los que se aplican las normas comunitarias en su contra, incluida la regla específica que aplica Facebook”. En este caso se incluye cualquier notificación a los usuarios para especificar la norma del estilo “blackface”, y también se deben enlazar los recursos mencionados anteriormente que explican el daño que quiere evitar esta norma. Facebook debe proporcionar una actualización detallada sobre su “estudio de viabilidad” de las anteriores recomendaciones del Consejo sobre este tema, incluida la naturaleza específica de cualquier limitación técnica y cómo se pueden superar.	E	◐	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META’S RESPONSE:** ● Completa, ◐ Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Preocupación punyabí sobre la RSS en India</b>				
1	Facebook debe traducir las normas comunitarias y las Normas de implementación internas al punyabí. Facebook debe proporcionar acceso a las Normas comunitarias en todos los idiomas de uso extendido que hablan los usuarios. De esta manera, estos entenderían bien las reglas que deben cumplir al usar los productos de Facebook. También les resultaría más fácil tratar con Facebook asuntos relativos a contenido que podría infringir sus derechos.	CP	●	■
2	En la misma línea que la recomendación del Consejo en el caso 2020-004-IG-UA, la empresa debe restaurar la revisión humana y el acceso a procesos de apelación humanos a los niveles previos a la pandemia lo antes posible y, al mismo tiempo, proteger la salud del personal y los contratistas de Facebook.	E	●	■
3	Facebook debe mejorar sus informes de transparencia para aumentar la información pública sobre las tasas de error; para ello, esta información puede estar disponible para cada país y lengua con normas comunitarias. El Consejo destaca que los informes de transparencia más detallados ayudarán al público a identificar áreas en que los errores son más comunes, incluidos los posibles impactos en grupos minoritarios, y podrán alertar a Facebook para que lo corrija.	T	◐	■
<b>Suspensión del expresidente Trump</b>				
1	El Consejo manifiesta que no siempre resulta útil establecer una distinción firme entre líderes políticos y otros usuarios influyentes. Es importante reconocer que otros usuarios con gran cantidad de público también pueden contribuir a que se produzcan daños graves. Deben aplicarse las mismas reglas a todos los usuarios de la plataforma, pero es importante el contexto al evaluar problemas de causalidad y la probabilidad y la inminencia de daños. Es importante el grado de influencia que un usuario tiene sobre los demás. (...) Facebook debe evaluar las publicaciones realizadas por usuarios influyentes en contexto de acuerdo con la forma en que podrían interpretarse, incluso si su mensaje provocador está formulado en un lenguaje diseñado para evitar responsabilidades; por ejemplo, mediante un mensaje superficial que insta a actuar de forma pacífica o legal. Facebook usó los seis factores contextuales del Plan de Acción de Rabat en este caso, y el Consejo considera que es una forma útil de evaluar los riesgos contextuales del lenguaje potencialmente dañino. El Consejo recalca que el tiempo es crucial es este tipo de situaciones; es necesario actuar antes de que los usuarios influyentes puedan causar un daño significativo y ello debe tener prioridad al interés periodístico y otros valores de comunicación política.	CP	◐	■
2	Cuando las publicaciones de usuarios influyentes suponen una alta probabilidad de daños inminentes, tal como se indica en las normas internacionales sobre derechos humanos, Facebook debe actuar y aplicar sus reglas de inmediato.	E	◐	■
3	Si bien todos los usuarios deben estar sujetos a las mismas políticas de contenido, hay factores exclusivos que se deben tener en cuenta al evaluar el discurso de líderes políticos. Los jefes de Estado y otros altos funcionarios del Gobierno tienen más poder para causar daño que otras personas. Facebook debe reconocer que las publicaciones que realizan los jefes de Estado y otros altos funcionarios del Gobierno pueden tener un mayor riesgo de fomentar, legitimar o incentivar la violencia, ya sea porque su alto nivel de confianza infunde mayor fuerza y credibilidad a sus palabras o porque sus seguidores pueden inferir que pueden actuar con impunidad. Al mismo tiempo, es importante proteger los derechos de las personas de escuchar discursos políticos. Sin embargo, si un jefe de Estado o alto funcionario de gobiernos publica reiteradamente mensajes que representan un riesgo de daño en virtud de las normas de derechos humanos internacionales, Facebook debe suspender la cuenta durante un plazo lo suficientemente largo como para proteger a las personas de daño inminente.	CP	●	■
4	Los períodos de suspensión deben durar lo suficiente para disuadir la mala conducta y pueden, en los casos correspondientes, incluir la eliminación de la cuenta o página.	CP	●	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ◐ Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

□ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Suspensión del expresidente Trump</b>				
5	Las restricciones sobre la libertad de expresión suelen imponerse por disposición u orden de agentes estatales poderosos contra personas que disienten y miembros de la oposición política. Facebook debe resistir la presión de los poderes públicos de silenciar a la oposición política. Al evaluar posibles riesgos, Facebook debe tener especial cuidado al considerar el contexto político relevante.	E	●	■
6	Cuando evalúa el discurso político de usuarios altamente influyentes, Facebook debe escalar de inmediato el proceso de moderación de contenido al personal especializado que esté familiarizado con el contexto lingüístico y político y que esté aislado de injerencias políticas, económicas y de influencias indebidas. Este análisis debe examinar la conducta de los usuarios altamente influyentes fuera de las plataformas de Facebook e Instagram para evaluar de forma adecuada todo el contexto relevante del lenguaje que podría ser dañino. Además, Facebook debe garantizar que dedica los recursos y aplica los conocimientos adecuados para evaluar los riesgos de daños que suponen las cuentas influyentes de todo el mundo.	E	✘	■
7	Facebook debe explicar de forma pública las reglas que usa al imponer sanciones en las cuentas de usuarios influyentes. Estas reglas deben garantizar que, cuando Facebook imponga una suspensión por tiempo limitado a la cuenta de un usuario influyente con el fin de reducir el riesgo de daños significativos, evaluará si ese riesgo disminuyó antes de que finalice la suspensión. Si Facebook determina que el usuario supone un riesgo grave de incitación a la violencia inminente, discriminación u otra acción ilegal en ese momento, se debe imponer otra suspensión por tiempo limitado cuando dicha medida sea necesaria para proteger la seguridad pública y proporcional al riesgo.	CP	●	■
8	Cuando Facebook implementa procedimientos especiales que se aplican a usuarios influyentes, estos deben estar bien documentados. No quedaba claro si Facebook aplicó diferentes normas en este caso, y el Consejo recibió muchos mensajes de preocupación sobre la posible aplicación de la concesión de interés periodístico. Es importante que Facebook aborde esta falta de transparencia y la confusión que causó.	T	●	■
9	Facebook debe proporcionar más información que ayude a los usuarios a entender y evaluar el proceso y los criterios para la aplicación de la concesión de interés periodístico. Facebook debe explicar con claridad cómo se aplica la concesión de interés periodístico a las cuentas influyentes, incluidas las de líderes políticos y otras figuras públicas.	CP	●	■
10	Respecto a la revisión de verificación cruzada, Facebook debe explicar con claridad los fundamentos, las normas y los procesos de revisión, incluidos los criterios para determinar qué páginas y cuentas se seleccionan para su inclusión.	CP	●	■
11	Facebook debe informar las tasas de error relativas y la coherencia temática de las determinaciones realizadas a través del proceso de comprobación en comparación con procedimientos habituales de aplicación de políticas.	T	✘	■
12	Cuando usuarios influyentes usen la plataforma de Facebook de forma indebida de una manera que cause graves efectos adversos sobre los derechos humanos, Facebook debe realizar una investigación exhaustiva del incidente. Facebook debe evaluar la influencia que tuvo y los cambios que podría implementar para identificar, evitar y mitigar los efectos adversos, así como explicar de qué manera los abordará en el futuro.	E	✘	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Suspensión del expresidente Trump</b>				
13	Facebook debe realizar una revisión integral de su posible contribución al discurso de fraude electoral y las tensiones exacerbadas que culminaron en los hechos de violencia del 6 de enero de 2021 en los Estados Unidos. Esta debe ser una reflexión abierta sobre el diseño y las decisiones normativas que tomó Facebook, que podrían permitir que su plataforma se utilice indebidamente. Facebook debe llevar a cabo este proceso con la debida diligencia, implementar un plan para actuar en función de sus resultados y comunicar abiertamente cómo va a abordar los efectos adversos sobre los derechos humanos con los que estuvo relacionado.	T	✘	
14	En los casos donde los usuarios de Facebook o Instagram puedan haber participado en crímenes atroces o infracciones graves de derechos humanos, así como en actividades de incitación en virtud del Artículo 20 del ICCPR, hay que tener en cuenta que la eliminación del contenido y la inhabilitación de cuentas, si bien reducen potencialmente el riesgo de daño, también pueden socavar las iniciativas de responsabilidad o incluso eliminar evidencias. Facebook tiene la responsabilidad de recopilar, preservar y, si corresponde, compartir información para colaborar en la investigación y el posible procesamiento judicial de infracciones graves de leyes penales, de derechos humanos y leyes humanitarias internacionales por parte de las autoridades competentes y los mecanismos de responsabilidad. La política corporativa de derechos humanos de Facebook debe aclarar los protocolos que la empresa aplica al respecto. La política también debe explicar cómo la información que estuvo previamente publicada en la plataforma puede ponerse a disposición de los investigadores que realizan su trabajo de acuerdo con las normas internacionales y las leyes aplicables de protección de datos.	T	●	
15	Facebook debe explicar en las normas comunitarias y pautas el proceso de faltas y penalizaciones para restringir perfiles, páginas, grupos y cuentas en Facebook e Instagram de una manera clara, integral y accesible. Estas políticas deben brindar a los usuarios suficiente información para entender cuándo se imponen las faltas (incluidas las excepciones o concesiones aplicables) y cómo se calculan las penalizaciones.	CP	●	
16	Facebook también debe proporcionar a los usuarios información accesible sobre la cantidad de infracciones, faltas y penalizaciones que se evaluaron en su contra, así como las consecuencias que habrá si se cometen más infracciones en el futuro.	E	●	
17	En sus informes de transparencia, Facebook debe incluir diversas restricciones de perfiles, páginas y cuentas, incluido el motivo y la manera en que se aplicaron las políticas, con información desglosada por región y país.	T	✘	
18	Facebook debe desarrollar y publicar una política que rijan su respuesta ante situaciones nuevas o de crisis en las que los procesos habituales no prevendrían ni evitarían el daño inminente. Si bien estas situaciones no siempre se pueden anticipar, los pasos de Facebook deben establecer los parámetros adecuados para esas acciones, incluido un requisito para revisar su decisión en un período fijo.	CP	●	
<b>Meme “dos botones”</b>				
1	Hacer arreglos técnicos que garanticen que el aviso que reciben los usuarios haga referencia a la norma comunitaria que aplica la empresa. Si Facebook determina que (i) el contenido no infringe la norma comunitaria que se indicó al usuario y (ii) que el contenido infringe una norma comunitaria diferente, el usuario debe recibir una notificación pertinente y otra oportunidad de apelación. Los usuarios deben contar siempre con la información correcta antes de acudir al Consejo.	E	●	
2	Incluir una excepción de sátira, que actualmente no se les comunica a los usuarios, en el contenido público de su norma comunitaria sobre lenguaje que incita al odio.	CP	●	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META’S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Meme “dos botones”</b>				
3	Cerciorarse de aplicar procedimientos adecuados para evaluar el contenido satírico y el contexto relevante según corresponda. Ello incluye proporcionar a los moderadores de contenidos: (i) acceso a equipos de operaciones locales de Facebook para recopilar información general y cultural relevante y (ii) tiempo suficiente para consultar con estos equipos y realizar la evaluación. Facebook debe garantizar que sus políticas para los moderadores de contenido sirvan como incentivo para realizar otras investigaciones o escalado en los casos en que el moderador no está seguro de que un meme constituya o no una sátira.	E	●	
4	Permitir a los usuarios indicar en su apelación si el contenido se incluye dentro de alguna de las excepciones de la política sobre lenguaje que incita al odio. Están incluidas las opciones de contenido satírico y si el usuario comparte contenido ofensivo a modo de repudio o para generar conciencia.	E	●	
5	Es necesario garantizar que las apelaciones basadas en excepciones a las políticas tengan prioridad en la revisión humana.	E	●	
<b>Protestas a favor de Navalny</b>				
1	Facebook debe enmendar y volver a redactar las normas comunitarias sobre bullying y acoso para así poder explicar la relación entre la Justificación de la política y las prohibiciones, así como las demás reglas que restringen contenido que sigue a estos casos.	CP	●	
2	Diferenciar entre bullying y acoso y proporcionar definiciones que distingan ambos actos. Además, la norma comunitaria debe explicar claramente a los usuarios la manera en que el bullying y el acoso difieren del lenguaje que solo ofende y puede estar protegido en virtud de leyes internacionales de derechos humanos.	CP	◐	
3	Es necesario definir claramente el enfoque en relación con distintas categorías de usuario objetivo y proporcionar ejemplos ilustrativos de cada categoría objetivo (por ejemplo, quiénes se consideran figuras públicas). Diseñar la norma comunitaria sobre bullying y acoso según las categorías de usuario incluidas actualmente en la política.	CP	●	
4	Incluir ejemplos ilustrativos de contenido que infringe la norma comunitaria sobre bullying y acoso y contenido que la cumple para clarificar los límites que establece la política y cómo estas distinciones pueden depender del tipo de identidad de la persona en cuestión.	CP	●	
5	Al evaluar contenido que incluya “afirmaciones negativas relacionadas con el carácter” contra un adulto no público, Facebook debe modificar la norma comunitaria para requerir un análisis del contexto sociopolítico del contenido. Facebook debe reconsiderar la aplicación de esta regla en debates políticos o públicos en los que la eliminación del contenido reprimiría el debate.	CP	✘	
6	Siempre que Facebook elimine contenido debido a una afirmación de carácter negativo y esta sea una sola palabra o frase en una publicación más amplia, debe notificar al usuario del hecho con prontitud para que este pueda volver a publicar el material sin la afirmación negativa.	E	●	
<b>Aislamiento de Öcalan</b>				
1	Restablecer de inmediato la pauta perdida de 2017 en las normas de implementación internas y en las preguntas frecuentes (el documento de referencia interno para moderadores de contenido), informar a todos los moderadores de contenido y poner en marcha formación sobre ello.	E	●	
2	Evaluar los procesos de moderación automatizados para la aplicación de la política sobre personas y organizaciones peligrosas. Si es necesario, Facebook debe actualizar sus clasificadores a fin de excluir datos de entrenamiento correspondientes a errores del sistema de cumplimiento de normas derivados de no haber implementado la pauta de 2017. Se deben añadir más datos de formación que ilustren el restablecimiento de esta pauta.	E	✘	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META’S RESPONSE:** ● Completa, ◐ Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Aislamiento de Ócalan</b>				
3	Publicar los resultados del proceso de análisis en curso para determinar si se ha extraviado alguna otra política. Esos resultados deben incluir descripciones de todas las políticas extraviadas, el período durante el cual permanecieron extraviadas y las medidas tomadas para restablecerlas.	T	●	
4	Reflejar en el “fundamento normativo” de la política sobre personas y organizaciones peligrosas que el respeto por los derechos humanos y la libertad de expresión (en particular, un debate abierto sobre abusos e infracciones de derechos humanos relacionados con el terrorismo y las iniciativas para erradicarlo) puede fomentar el valor de “seguridad” y que es importante que la plataforma ofrezca un espacio para que se entablen estos debates. Si bien puede haber ocasiones en las que sea complicado mantener un equilibrio entre los valores de “seguridad” y “expresión”, el fundamento normativo debe especificar en mayor detalle qué “daños en el mundo real” busca evitar y no aplicar la política cuando se reprime el derecho a la expresión.	CP	●	
5	Añadir a la política sobre personas y organizaciones peligrosas una explicación clara de lo que excluye el término “apoyo”. Los usuarios deben poder debatir supuestas infracciones o abusos de derechos humanos sufridos por miembros de las organizaciones designadas. Ello no debe limitarse a personas detenidas. Se debe poder debatir sobre los derechos protegidos por las convenciones de derechos humanos de la ONU, según se cita en la Política corporativa sobre derechos humanos de Facebook. Se deben permitir, por ejemplo, debates sobre denuncias de tortura o tratamiento o castigo cruel, inhumano o degradante; infracciones del derecho a un juicio justo; ejecuciones extrajudiciales, sumarísimas o arbitrarias; desaparición forzosa; o la revocación extraordinaria de la ciudadanía que dejan a una persona en condición de apátrida. También deben protegerse las peticiones de responsabilidad por infracciones y abusos de los derechos humanos. Se debe excluir de la protección el contenido que incita a actos violentos o recluta a personas para que se unan o, de cualquier otro modo, den apoyo material a organizaciones que Facebook haya designado, incluso si en ese contenido se analizan también cuestiones relativas a los derechos humanos. La intención del usuario, el contexto general en el que hace la publicación y la interpretación de ella que hacen otros usuarios son factores clave para determinar la probabilidad de daños reales que podría ocasionar dicha publicación.	CP	●	
6	Explicar en las normas comunitarias cómo los usuarios pueden dejar clara la intención de su publicación ante Facebook. A esto contribuiría la implementación de la recomendación existente del Consejo de divulgar públicamente la lista de personas y organizaciones designadas de la empresa (se puede consultar el caso 2020-005-FB-UA). Facebook debe proporcionar ejemplos ilustrativos sobre dónde se sitúa la línea divisoria entre el contenido prohibido y el admitido, incluido aquel que está relacionado con la aplicación de la regla que aclara lo que se excluye del término “apoyo”.	CP	●	
7	Garantizar la participación de partes interesadas pertinentes en el cambio propuesto para la política. Este proceso debe realizarse mediante el Foro de políticas de productos de Facebook y debe incluir una solicitud pública de sugerencias. Facebook debe propiciar esta participación en varios idiomas y en todas las regiones para garantizar que se forje un diálogo efectivo con las personas más afectadas por los daños que esta política intenta evitar. También se debe incluir a organizaciones académicas, de derechos humanos y de la sociedad civil con conocimiento especializado en esos daños, así como en los perjuicios que puedan producirse debido a una aplicación excesiva de la política actual.	CP	●	
8	Es necesario garantizar que los moderadores de contenidos reciban formación y asesoramiento interno respecto a cualquier política nueva. Los moderadores de contenido deben recibir los recursos adecuados para poder entender la nueva política y contar con el tiempo suficiente para tomar decisiones al aplicarla.	E	●	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

□ Recomendación ignorada, rechazada o redefinida



Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Aislamiento de Ócalan</b>				
9	Garantizar que los usuarios reciban una notificación cuando se elimine contenidos suyos. La notificación ha de indicar si la eliminación se debe a una solicitud de poderes públicos o a una infracción de las normas comunitarias, o si es consecuencia de una declaración de poderes públicos en la que se afirma que infringe una ley nacional (y el alcance jurisdiccional de esa eliminación).	E	●	■
10	Aclarar a los usuarios de Instagram que las normas comunitarias de Facebook se aplican a Instagram del mismo modo en que se aplican a Facebook, de conformidad con la recomendación del caso 2020-004-IG-UA.	CP	●	■
11	Incluir información sobre la cantidad de solicitudes de eliminación de contenidos recibidas por Facebook que han presentado poderes públicos argumentando infracciones de las normas comunitarias (frente a infracciones de legislación nacional), y el resultado de esas solicitudes.	T	◐	■
12	Incluir información más integral, desglosada por región e idioma, sobre las tasas de error en la aplicación de las reglas sobre “exaltación” y “apoyo” de personas y organizaciones peligrosas.	T	◐	■
<b>Bot de Myanmar</b>				
1	Facebook debe garantizar que sus normas de implementación internas estén disponibles en el idioma en que los moderadores de contenidos revisan los contenidos. En caso de tener que priorizar, Facebook debe centrarse primero en los contextos en los que los derechos humanos corran un mayor riesgo.	E	◐	■
<b>Confinamientos de el COVID en Brasil</b>				
1	Facebook debe llevar a cabo un análisis de proporcionalidad para identificar un conjunto de medidas menos invasivas que la eliminación del contenido. Cuando sea necesario, se deben usar medidas menos invasivas en los casos en que el contenido relacionado con el COVID-19 distorsione los consejos de autoridades sanitarias internacionales y cuando se identifique un posible daño físico que no sea inminente. Las medidas recomendadas incluyen las siguientes: (a) marcar el contenido para alertar a los usuarios sobre la naturaleza controvertida del contenido de una publicación y proporcionar enlaces a los puntos de vista de la Organización Mundial de la Salud y autoridades sanitarias nacionales; (b) crear obstáculos en las publicaciones para evitar que los usuarios interactúen con ellas o las compartan, y (c) bajar la calificación para disminuir la visibilidad en las secciones de noticias de otros usuarios. Todas estas medidas de cumplimiento de normas deben comunicarse con claridad a todos los usuarios y estar sujetas a apelación.	E	✘	■
2	Dado el contexto de la pandemia de COVID-19, Facebook debe realizar ajustes técnicos para priorizar la verificación de datos que compartan las autoridades públicas y que llamen la atención de la empresa, para evitar así la posible información errónea sobre salud; además es necesario tener en cuenta el contexto local.	E	✘	■
3	Facebook debe ofrecer mayor transparencia en la política sobre noticias falsas de las normas comunitarias respecto a cuándo el contenido cumple los requisitos de verificación de información, lo que incluye saber si las cuentas de instituciones públicas están sujetas a este tipo de comprobación.	CP	●	■
<b>Publicación compartida de Al Jazeera</b>				
1	Añadir criterios y ejemplos ilustrativos a su política sobre personas y organizaciones peligrosas para mejorar la comprensión de las excepciones relativas al debate neutral, las actitudes de condena y los fines periodísticos.	CP	●	■
2	Garantizar que las actualizaciones de las normas comunitarias se traduzcan con celeridad a todos los idiomas disponibles.	CP	◐	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META’S RESPONSE:** ● Completa, ◐ Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

◻ Progreso realizado

◻ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Publicación compartida de Al Jazeera</b>				
3	Contratar a una entidad independiente no asociada con ningún bando del conflicto entre Israel y Palestina para efectuar un examen exhaustivo a fin de determinar si la moderación de contenido de Facebook en árabe y en hebreo, incluido el uso de la automatización, se ha aplicado sin sesgos. Este análisis no debe centrarse únicamente en el trato concedido al contenido palestino o propalestino, sino que también debe revisar cómo se trata el contenido que incita a la violencia contra cualquier objetivo potencial, con independencia de su nacionalidad, origen étnico, religión, creencias u opiniones políticas. La revisión debe analizar el contenido que hayan publicado los usuarios de Facebook ubicados en Israel y los Territorios Palestinos Ocupados, así como fuera de esos ámbitos geográficos. Tanto el informe como sus conclusiones deben hacerse públicos.	T	●	
4	Formalizar un proceso transparente en el que se detalle cómo Facebook recibe solicitudes de poderes públicos de eliminación de contenido y les da respuesta, y garantizar que estas se incluyan en los informes de transparencia. Los informes de transparencia deben distinguir entre las solicitudes de poderes públicos que han derivado en eliminaciones de contenidos debidas a infracciones de las normas comunitarias y las solicitudes que han llevado a eliminaciones o bloqueos geográficos de contenidos por infringir la legislación local, además de las solicitudes que no han tenido más consecuencias.	T	●	
<b>Protestas en Colombia</b>				
1	Publicar ejemplos ilustrativos de la lista de insultos designados como infractores en las normas comunitarias sobre lenguaje que incita al odio, incluidos casos dudosos con términos que podrían ser dañinos en ciertos contextos, pero no en otros.	CP	●	
2	Vincular la breve explicación de la concesión de interés periodístico provista en la introducción de las normas comunitarias a la argumentación más detallada del Centro de transparencia de Facebook, en la que se explica cómo se aplica esta política. La empresa debe complementar esta explicación con ejemplos ilustrativos correspondientes a diversos contextos, incluida la cobertura periodística de manifestaciones masivas.	CP	●	
3	Publicar y divulgar criterios claros que permitan a los revisores de contenidos derivar el material de interés público para una segunda revisión en el caso de que pueda infringir las normas comunitarias, pero que también cumpla los requisitos para adquirir la concesión de interés periodístico. Esos criterios deben abarcar el contenido en que se muestren grandes manifestaciones sobre asuntos políticos.	E	✘	
4	Informar a todos los usuarios que denuncien contenidos considerados como infractor en caso de que se aplique la concesión de interés periodístico a tal publicación y esta se mantenga en la plataforma por motivos de interés público. En la correspondiente notificación, se debe incluir un enlace a la explicación de esa concesión de interés periodístico en el Centro de transparencia.	E	●	
<b>Insultos en Sudáfrica</b>				
1	Es necesario informar a los usuarios que han infringido la regla específica incluida en las normas comunitarias sobre lenguaje que incita al odio en el idioma en el que usan Facebook, según la recomendación en la decisión del caso 2020-003-FB-UA (Armenios en Azerbaiyán) y del caso 2021-002-FB-UA (Representación de Zwarte Piet). En este caso concreto, se debe notificar al usuario que ha infringido la prohibición de emplear insultos. El Consejo también señaló la respuesta de Facebook a la recomendación núm. 2 en la decisión del caso 2021-002-FB-UA, que describe un nuevo clasificador con el que debe poder notificar a los usuarios de Facebook que hablan inglés que su contenido ha infringido la regla sobre insultos. El Consejo espera que Facebook proporcione información que confirme la implementación de este clasificador en el caso de los usuarios angloparlantes, así como datos sobre el plazo para realizar esa implementación en el caso de los usuarios que hablan otros idiomas.	E	●	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

□ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Cinturón Wampum</b>				
1	Notificar a los usuarios de forma oportuna y precisa toda acción que la empresa realice respecto al contenido relacionado con su apelación. Cuando corresponda (incluidos los casos de errores del sistema de cumplimiento de normas, como este), en el aviso enviado al usuario se debe reconocer que la acción se ha realizado como consecuencia del proceso de revisión del Consejo asesor de contenidos. Meta debe compartir los mensajes enviados a los usuarios cuando las acciones del Consejo afecten a las decisiones de contenido que los usuarios apelaron, a fin de demostrar su cumplimiento con esta recomendación. Estas acciones deben realizarse en relación con todos los casos corregidos en la etapa de selección del proceso del Consejo.	E	●	
2	Estudiar cómo afectarían a la precisión y la productividad de los revisores determinadas modificaciones de los métodos de revisión secundaria. En particular, el Consejo solicita una evaluación de los índices de precisión cuando se informa a los moderadores de contenido que realizan una revisión secundaria, de modo que sepan que la resolución inicial se revocó. Idealmente, este experimento debe ofrecer la posibilidad para los usuarios de proporcionar contexto relevante que permita a los revisores evaluar sus contenidos, en consonancia con las recomendaciones previas del Consejo. Meta debe compartir los resultados de estas evaluaciones sobre la precisión con el Consejo y resumir los resultados en el informe de transparencia que presenta ante el Consejo de forma trimestral para demostrar que sigue esa recomendación.	E	●	
3	Llevar a cabo evaluaciones de precisión centradas en las concesiones correspondientes a las políticas sobre lenguaje que incita al odio, que amparan la expresión artística y la expresión sobre violaciones de los derechos humanos (por ejemplo, actos de condena, concienciación, uso autorreferencial o uso inspirador). En esta evaluación también se debe investigar específicamente cómo el lugar de residencia del revisor afecta a la capacidad de los moderadores de evaluar con precisión el lenguaje que incita al odio y las respuestas a contenidos ofensivos originados en la misma región u otras regiones. El Consejo entiende que este análisis probablemente requiera el desarrollo de ejemplos apropiados y etiquetados con precisión del contenido relevante. Meta debe compartir los resultados de esta evaluación con el Consejo; entre otras cosas, debemos saber cómo se usarán esos resultados para mejorar las operaciones de cumplimiento de normas y el desarrollo de políticas, y si se pretende llevar a cabo evaluaciones de la precisión de los revisores de forma regular en torno a estas concesiones y resumir los resultados en el Informe de transparencia trimestral del Consejo para demostrar que, en efecto, ha cumplido esta recomendación.	E	●	
<b>Bebida de ayahuasca</b>				
1	El Consejo reitera la recomendación de la decisión del caso 2020-004-IG-UA y la decisión del caso 2021-006-IG-UA de que Meta debe explicar a los usuarios que aplica las normas comunitarias de Facebook en Instagram, con varias excepciones concretas. El Consejo señala la respuesta de Meta a estas recomendaciones. Si bien es posible que Meta esté tomando otras medidas para cumplir con las recomendaciones, el Consejo recomienda que la empresa actualice la introducción a las normas comunitarias de Instagram (el resumen de las normas comunitarias) en un plazo de 90 días para informar a los usuarios de que, si se considera que un contenido infringe las normas en Facebook, también se considerará que infringe las normas en Instagram, tal como se afirma en el Centro de transparencia de la empresa, con algunas excepciones.	E	●	
2	El Consejo reitera su recomendación de la decisión del caso 2021-005-FB-UA y la decisión del caso 2020-005-FB-UA de que Meta debe explicar a los usuarios qué regla concreta de la política sobre contenidos han infringido.	E	●	

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

■ Implementación demostrada con información publicada

□ Progreso realizado

□ Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación

■ Recomendación ignorada, rechazada o redefinida

Identificador de la recomendación	Recomendación del Consejo asesor de contenido	Categoría	Evaluación de la	
			Respuesta de Meta del Consejo	Implementación del Consejo
<b>Bebida de ayahuasca</b>				
3	Con el fin de respetar las diversas expresiones y prácticas tradicionales y de carácter religioso, el Consejo recomienda que Meta modifique las normas comunitarias de Instagram y la norma comunitaria de Facebook sobre bienes regulados para permitir el debate positivo sobre los usos tradicionales y religiosos de las drogas sin fines medicinales cuando haya evidencias históricas de tales usos. Asimismo, el Consejo recomienda que Meta haga públicas todas las concesiones, incluidas las actuales.	CP	●	■
<b>Presuntos delitos en Raya Kobo</b>				
1	Meta debe reformular el valor de "seguridad" para reflejar que el lenguaje que incita al odio en internet puede suponer un riesgo para la seguridad física de las personas y el derecho a la vida, además de los riesgos de intimidación, exclusión y silenciamiento.	CP	●	■
2	Las normas comunitarias de Facebook deben reflejar que, en un contexto de guerra y conflictos violentos, los rumores no verificables suponen un mayor riesgo para los derechos a la vida y la seguridad de los demás. Ello debe quedar reflejado en todos los niveles del proceso de moderación.	CP	✘	■
3	Meta debe solicitar con la debida diligencia una evaluación independiente sobre los derechos humanos a fin de determinar cómo se usan Facebook e Instagram para difundir lenguaje que incita al odio y rumores infundados que aumentan el riesgo de violencia en Etiopía. En esta evaluación, se deberá revisar la eficacia de las medidas que Meta ha tomado para evitar el uso indebido de sus productos y servicios en Etiopía, así como de las que ha adoptado para hacer posible la elaboración de informes corroborados de interés público sobre atrocidades que atentan contra los derechos humanos en Etiopía. En esta evaluación también se deberán revisar las funciones de Meta relacionadas con el idioma en Etiopía y si estas son lo bastante eficaces para proteger los derechos de los usuarios. La evaluación deberá abarcar el período comprendido entre el 1 de junio de 2020 y el presente. La empresa deberá completar la evaluación en un plazo de seis meses a partir del momento en que responda a estas recomendaciones. La evaluación deberá publicarse en su totalidad.	T	●	■

**CATEGORÍA:** E– Ejecución, T–Transparencia, CP–Política sobre Contenidos

**META'S RESPONSE:** ● Completa, ● Bastante completa, ✘ No completa

**IMPLEMENTATION**

- Implementación demostrada con información publicada
- Progreso realizado

- Meta informó de que ya estaba implementado esta opción o indicó que ya lo estaba haciendo, pero no publicó información que demostrara esa implementación
- Recomendación ignorada, rechazada o redefinida



