



Oversight Board
Q3 2022
transparency report



Glossary of terms

Annual report – A report published each year by the Oversight Board that provides a summary of the cases it selects and reviews, as well as an overview of its operations.

Bylaws – [These](#) specify the Oversight Board’s operational procedures.

Case Management Tool (CMT) – The platform created by Meta and used by the Oversight Board to receive and review case submissions, and collect and store case files.

Case Selection Committee – A sub-committee of the Board, formed to address case selection.

Case Selection Team – A team within the Oversight Board Administration that assists the Case Selection Committee with identifying cases for panel review.

Facebook content policies – Facebook and Instagram’s content policies and procedures that govern content on the platforms (e.g., Community Standards and/or Community Guidelines).

Meta’s legal review – Step in case selection process where Meta may exclude cases from the shortlist which are ineligible for review by the Board in accordance with the Bylaws. More detail about this stage can be found in the [Rulebook for Case Review and Policy Guidance](#) (page 8).

Meta-referred case – A case submitted to the Oversight Board by Meta. Meta has the ability to expedite cases for review.

Longlist – An initial list of cases drawn up by the Case Selection Team. This is based on selection criteria set out by the Case Selection Committee.

Oversight Board Administration – The full-time professional staff that support Board Members and the day-to-day operations of the Board.

Panel – Five Members of the Oversight Board assigned to review a case.

Policy advisory statement – A statement appended to an Oversight Board decision on a specific case that reflects policy considerations beyond the binding content decision.

Shortlist – A small number of cases chosen from the longlist by the Case Selection Committee to be considered for selection.

User appeal – An appeal submitted by a Facebook or Instagram user to the Oversight Board for review.



Transparency Report for third quarter of 2022

This transparency report for the third quarter of 2022 (July 1 – September 30, 2022) sets out key statistics on cases selected by the Board, as well as the decisions and recommendations we made in this quarter.

In this quarter, the Board published two case decisions, “Mention of the Taliban in news reporting” and “Colombian police cartoon.”



Key Findings – Third quarter 2022¹



~270,848 cases submitted to the Oversight Board, of which five were submitted by Meta.

43 user-submitted cases longlisted.

20 user-submitted cases shortlisted.

Announced that six cases had been selected, four submitted by users and two referred by Meta, as well as a request for a policy advisory opinion on the removal of COVID-19 misinformation.

Two case decisions were published in this period. These contained 10 recommendations for Meta.

¹ Cases may pass through stages of the review process in multiple reporting periods. The data presented reflect the number of cases that are within each stage of the review process during the reporting period in question. Thus, a case submitted in Q2 but longlisted in Q3 would appear in “submitted cases” in Q2 and “longlisted cases” in Q3.



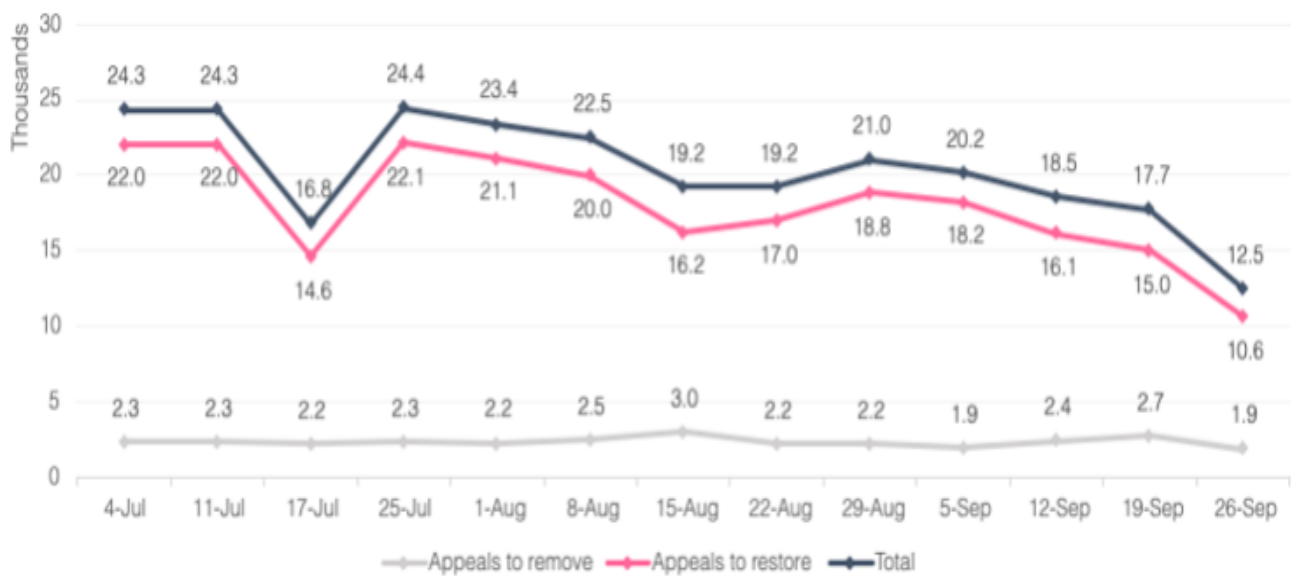
Q3 2022 Submitted User Cases

Where users have exhausted Meta’s appeals process, they can challenge the company’s decision by appealing eligible content to the Oversight Board.

In Q3 2022, **270, 843 cases** were submitted by users, a reduction of 22% on Q2 2022. In total, users submitted approximately two and a quarter million cases to the Board from October 2020 to September 2022.

Estimated number of cases submitted to Oversight Board by week

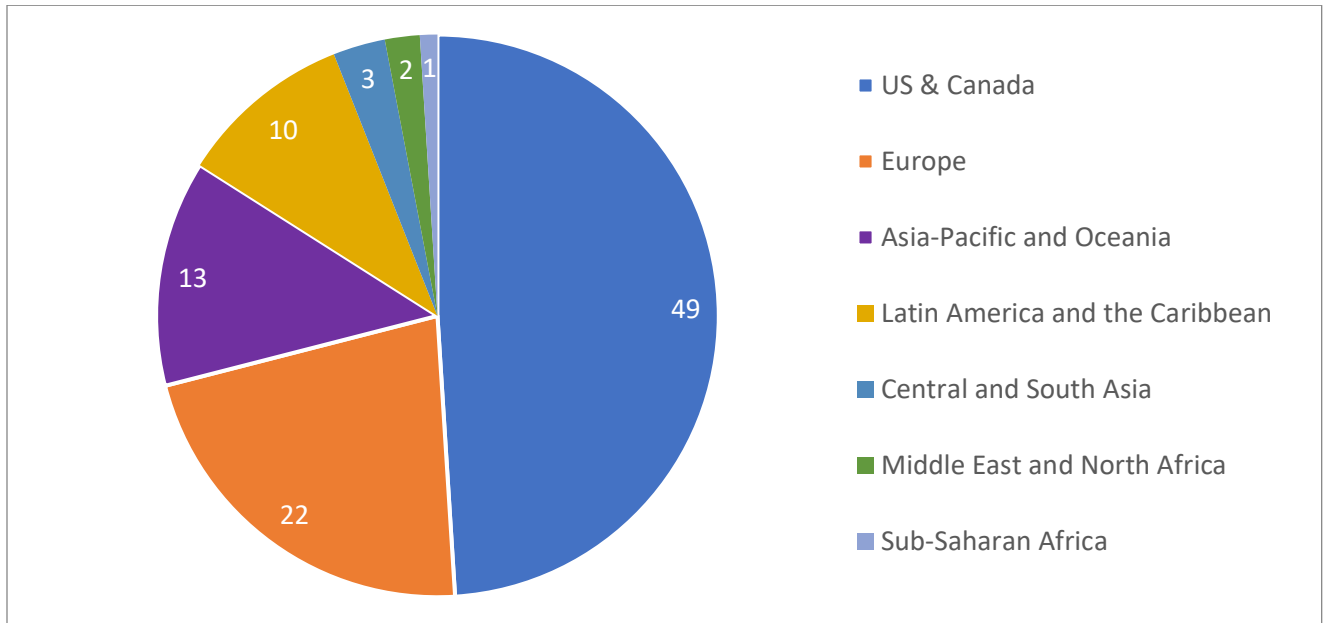
Number of cases, thousands



The Board received the highest number of appeals in Q3 in the week commencing 25 July, with users submitting around 24,400 cases over seven days. From Friday 30 September – Sunday 2 October, a bug in the appeals system, that has since been fixed, led to a delay in the delivery of cases to the board, resulting in a significant decline in appeals for that week. As with previous quarters, the vast majority (89%) of cases submitted to the Board were appeals to restore content, while 11% of cases were appeals to remove other users’ content.

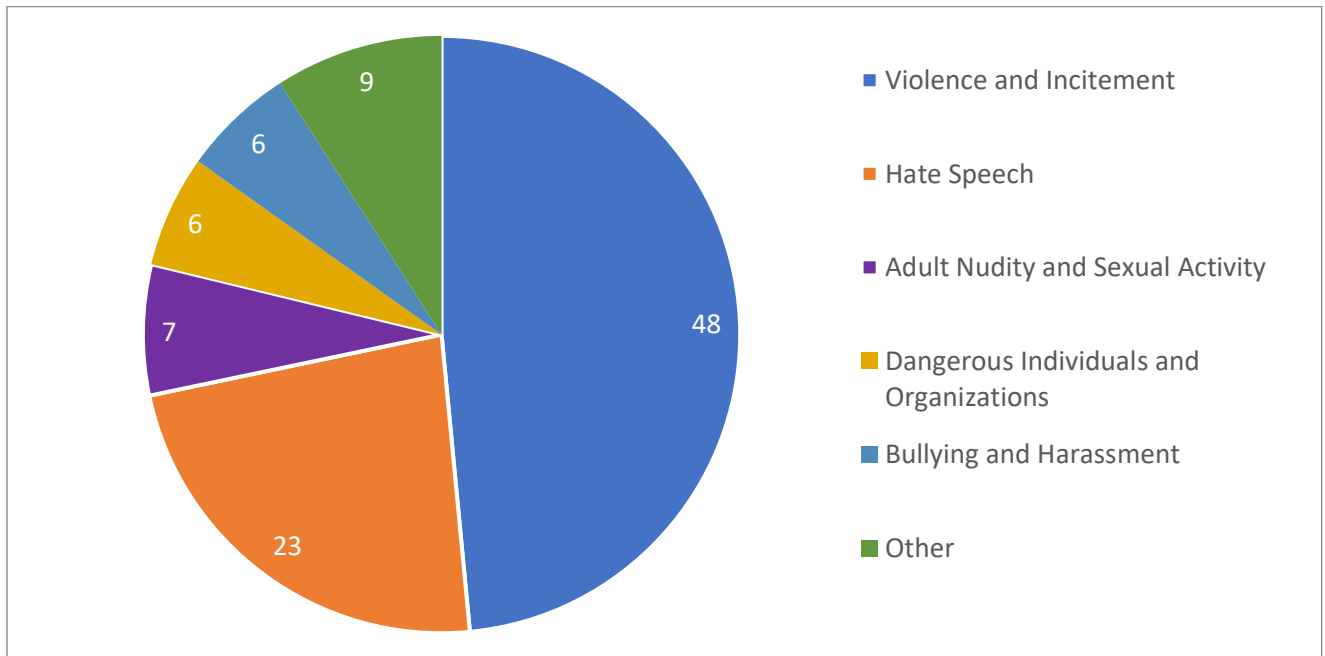


Estimated cases submitted by user-selected region (Percent)



Around half (49%) of submitted cases came from the United States & Canada, followed by 22% from Europe, 13% from Asia Pacific and Oceania, and 10% from Latin America and the Caribbean.

Estimated appeals to *restore* content to Facebook/Instagram by Community Standard (Percent)



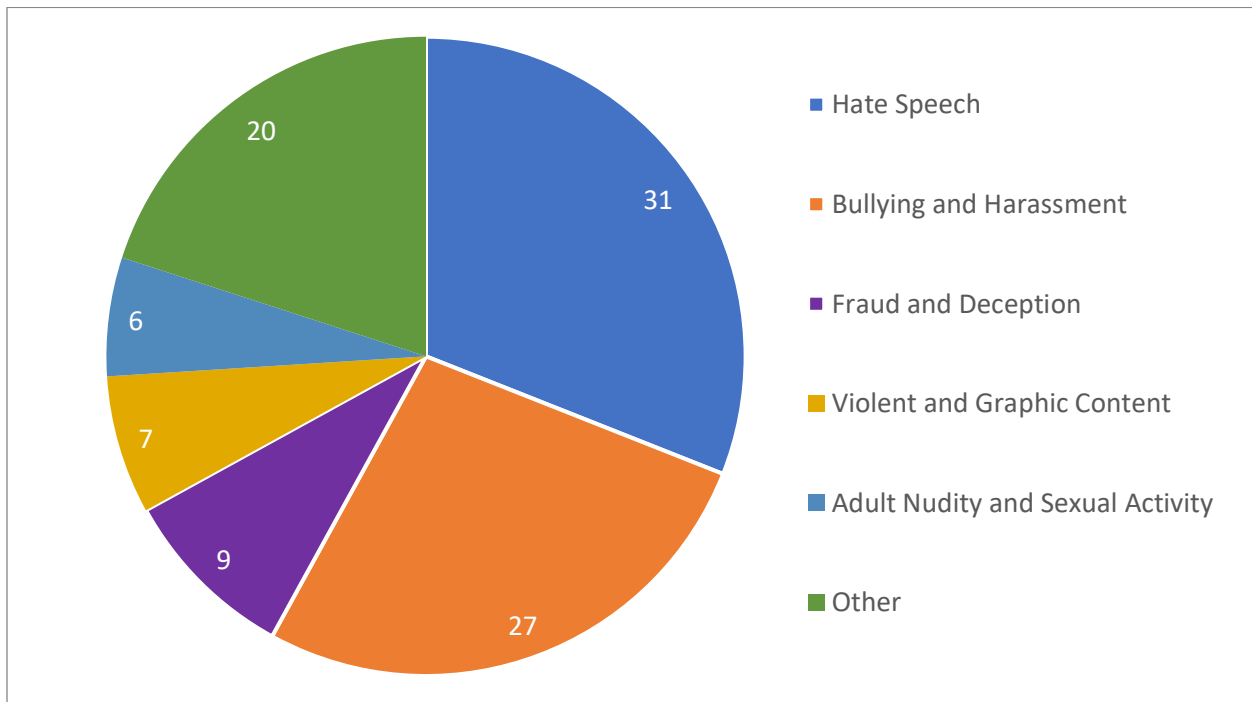


Users primarily submitted appeals to restore content which Meta removed for violating its policies on Violence and Incitement (48%), Hate Speech (23%) and Adult Nudity and Sexual Activity (7%).

The share of cases relating to content removed for supposedly violating Meta’s Violence and Incitement policy continued to represent around half of appeals to the Board to restore content, representing 51% in Q2 2022 and 48% in Q3 2022.

In Q3 2022, the share of cases about content removed for supposedly violating Meta’s Bullying and Harassment policy reached an all-time low of 6%. In Q3 2021, by contrast, cases about content removed under Meta’s Bullying and Harassment policy made up more than a third (34%) of appeals to restore content – more than any other Community Standard.

Estimated appeals to *remove* content from Facebook/Instagram by user-selected Community Standard (Percent)



For user appeals to *remove* content from Facebook or Instagram, the Community Standard which has supposedly been violated is determined not by Meta (as with appeals to *restore* content) but by users themselves.

As shown in the chart above, in Q3 2022 users submitted the most appeals to remove posts they thought violated Meta’s Hate Speech Community Standard (31%), followed by Bullying and Harassment (27%), and Fraud and Deception (9%).

Cases submitted by platform

Percent

Facebook

92%



Instagram

8%

The vast majority of cases submitted by users (92%) concerned content shared on Facebook, with only 8% of cases concerning content shared on Instagram.

While cases about content on Facebook still vastly outnumber cases about content on Instagram, 8% represents, by far, the largest share of cases about posts on Instagram since the Board first shared this data point in Q2 2021. Previously, the largest share of appeals from Instagram had been just 2.5% (in Q2 2022).

Q3 2022 Longlisted User Cases

In this period, **43 user-submitted cases were longlisted** for the Case Selection Committee’s consideration. These were chosen using the selection criteria set by the committee, including the ‘overarching criteria for case selection’ available [here](#).

These cases covered 33 pieces of content on Facebook and 10 pieces of content on Instagram. 24 were cases to restore content and 19 were cases to remove content.

Longlisted cases by region and affected country²

Number of cases

Asia Pacific and Oceania	3
Philippines	2
China	1
Europe	13
Ukraine	5
Russia	3
Romania	1
Italy	1
United Kingdom	2
Netherlands	1
United States & Canada	5
United States	5
Sub-Saharan Africa	8
Ethiopia	1
Nigeria	1
Democratic Republic of the Congo	1
Guinea-Bissau	1
Somalia	2
Kenya	1

² 'Countries Affected' is a user-selected field in the Board’s appeal process and users can select multiple countries. For this reason, 43 user-submitted cases were longlisted but a tally of 52 is listed here. While the user selects the relevant country in the first instance, the Case Selection Team also have the ability to change the country to improve accuracy.



Namibia	1
Central & South Asia	6
Afghanistan	2
Sri Lanka	1
Thailand	1
Pakistan	2
Latin America and Caribbean	5
Brazil	4
Colombia	1
Middle East and North Africa	12
Iran	1
Israel	5
Egypt	2
Palestine	3
Jordan	1
Total	52

Of the 43 cases longlisted by the Board in this quarter, 19 concerned content which had been left up on Facebook or Instagram and had not been deemed to have violated Meta’s rules. As such, they have not been assigned a Community Standard below, and are included under “undefined.” Of those cases where the content had been deemed by Meta to violate its rules, the most common Community Standards were Dangerous Individuals and Organizations (eight cases), followed by Hate Speech (four cases), Adult Nudity and Sexual Activity (three cases) and Violent and Graphic Content (three cases).

Longlisted cases by Community Standard

Number of cases	
Undefined ³	19
Dangerous Individuals and Organizations	8
Hate Speech	4
Adult Nudity and Sexual Activity	3
Violent and Graphic Content	3
Other	6
Total	43

Q3 2022 Shortlisted User Cases

The Case Selection Committee identifies a shortlist of cases from the longlist to consider for selection, selecting cases by majority vote.

In this quarter, the Case Selection Committee reviewed and shortlisted cases on four occasions, **shortlisting 20 cases in total.**

³ For content that is still live on Facebook and Instagram and reported by users, the applicable Community Standard violated is undefined as no Community Standard is purportedly violated.



The shortlist is sent to Meta’s legal team to review for eligibility as, per the Bylaws, certain cases are not eligible for review by the Board due to legal restrictions. While this eligibility review does not include a re-review on the merits, in practice, Meta has also assessed whether its original decision on a piece of content was correct or not. In many cases, it then decides to remove content which it incorrectly left up on its platforms or restore content which was incorrectly removed. This does not affect the eligibility of the content under review.

Of the 20 cases shortlisted in this period, 19 were confirmed as eligible by Meta’s legal team. Meta found that one shortlisted case relating to Thailand was ineligible, citing legal obligations in the country (see Annex I, point 12). Meta determined that its original decision on the piece of content was incorrect in 12 out of the 20 cases shortlisted by the Board. In 11 of these cases, Meta found it had incorrectly removed content and restored it. In one of case, it found it had wrongly left up content and removed it.

While this is only a small sample, and the Board intentionally seeks out difficult and challenging cases, it is noted that Meta found its original decision to have been incorrect in 60% of cases the Board shortlisted in Q3 2022. This is the same as the figure in Q2 2022. In Q1 2022 Meta found its original decision to be incorrect in 70% of cases the Board shortlisted. The Board continues to raise with Meta the questions this poses for the accuracy of the company's content moderation and the appeals process the company applies before cases reach the Board.

Cases where Meta identified that its original decision on content was incorrect

Number of cases

<i>Community Standard</i>	<i>Facebook</i>	<i>Instagram</i>	<i>Total</i>
Hate Speech	3	1	4
Dangerous Individuals and Organizations	2	1	3
Violence and Incitement	1	0	1
Adult Sexual Exploitation	1	0	1
Bullying and Harassment	1	0	1
Violent and Graphic Content	1	0	1
Undefined	1	0	1
Total	10	2	12



Q3 2022 Cases Submitted by Meta

In addition to appeals from users, Meta can also refer significant and difficult cases to the Board for consideration.

Meta submitted five cases to the Oversight Board during this period. Meta removed the content in three cases, which were related to the Violence and Incitement Community Standard, the Adult Nudity and Sexual Activity Community Standard and the Restricted Goods and Services Community Standard (one case each). The remaining two cases concerned content that the company had left up on Facebook.

Meta referrals submitted⁴

<i>Case ID</i>	<i>Name</i>	<i>Meta's decision</i>	<i>Platform</i>	<i>Language</i>	<i>Community Standard</i>	<i>Countries⁵</i>
2022-014-FB-MR	Sri Lanka pharmaceuticals	Leave up	FB	English	N/A	Sri Lanka
N/A	N/A	Take down	IG	English	Violence and Incitement	US
N/A	N/A	Take down	FB	Italian	Adult Nudity and Sexual Activity	Italy
N/A	N/A	Take down	IG	English	Restricted Goods and Services	US
N/A	N/A	Leave up	FB	English	N/A	India

Policy advisory opinion on removal of COVID-19 misinformation (PAO 2022-01)

On 26 July, the Oversight Board announced it had accepted a request from Meta for a policy advisory opinion on the removal of COVID-19 misinformation (PAO 2022-01). The company's current approach to removing COVID-19 misinformation is outlined in its policy on harmful health misinformation. In its request to the Board, Meta asked whether it should continue to remove content or whether another, less restrictive, approach would better align with the company's values and human rights responsibilities.

⁴ Cases which are not selected for assignment do not have a Case ID.

⁵ Countries listed do not necessarily align with countries assigned in longlisted cases above as a more thorough review is done at this stage of the appeals process to identify the principal countries concerned.



Meta asked the Board to consider the following policy options: continue removing certain COVID-19 misinformation; stop removing COVID-19 misinformation and instead reduce the distribution of the claims; third-party fact-checking, whereby content currently subject to removal would be sent to independent third-party fact-checkers for evaluation; adding labels to content that would provide links to authoritative information. The Board's recommendations and policy advisory opinion may not be limited to these options.

The Board invited public comments on the case between July 26 and August 25, 2022 and will publish its decision in 2023.

Q3 2022 Selected Cases

The Case Selection Committee selects cases for review, which are then announced publicly on the Board's website. In Q3, for the first time, the Board announced it would address two cases together (2022-009-IG-UA and 2022-010-IG-UA).

Cases selected

<i>Case ID</i>	<i>Name</i>	<i>Date announced</i>	<i>Platform</i>	<i>Source</i>	<i>Community Standard</i>	<i>Countries</i>
2022-007-IG-MR	UK drill music	7/26/22	IG	Meta referral	Violence and Incitement	UK
2022-008-FB-UA	Russian poem	7/26/22	FB	User (appeal to restore content)	Hate Speech & Violent and Graphic Content	Latvia, Russia, & Ukraine.
2022-009-IG-UA	Gender identity and nudity	7/26/22	IG	User (appeal to restore content)	Sexual Solicitation	US
2022-010-IG-UA	Gender identity and nudity ⁶	7/26/22	IG	User (appeal to restore content)	Sexual Solicitation	US
2022-011-IG-UA	Video after Nigeria church attack	9/15/22	IG	User appeal (to restore content)	Violent and Graphic Content	Nigeria

⁶ This is the first time the Board is addressing two cases together, 2022-009-IG-UA and 2022-010-IG-UA. They therefore have the same case name.



2022-012- IG-MR	India sexual harassment video	9/15/22	IG	Meta referral	Adult Sexual Exploitation	India
--------------------	-------------------------------------	---------	----	---------------	------------------------------	-------



Q3 2022 Published Case Decisions

After being selected, the Board assigns cases to a five-member panel. Members of the panel include at least one member from the region implicated in the content and a mix of gender representation. The panel looks at whether Meta’s decision is consistent with Facebook’s content policies and values, and its international human rights responsibilities. The Board’s decisions are binding, and Meta must implement them within seven days of publication.

In this period, the Oversight Board issued two decisions. It overturned Meta’s original decision in both of them.

Cases decided

<i>Case ID</i>	<i>Name</i>	<i>Platform</i>	<i>Source</i>	<i>Language of content</i>	<i>Community standard</i>	<i>Countries⁷</i>	<i>Outcome</i>
2022-004-FB-UA	Colombian police cartoon	FB	User appeal (to restore content)	Spanish	Dangerous Individuals and Organizations & Violence and Incitement	Colombia	Overturned
2022-005-FB-UA	Mention of the Taliban in news reporting	FB	User appeal (to restore content)	Urdu	Dangerous Individuals and Organizations	India & Afghanistan	Overturned

Human Rights standards referenced in decisions

In making its decisions the Oversight Board considers international human rights standards. The table below shows which human rights standards have been referenced in decisions published this quarter.

⁷ Countries listed do not necessarily align with countries identified in the metadata of longlisted cases above as a more thorough review is done at this stage of the appeals process.



Human rights standards referenced

<i>Source</i>	2022-004-FB-UA	2022-005-FB-UA
---------------	----------------	----------------

UN Treaties

ICCPR

Equality & Non-Discrimination (Art. 2 & 26)		✓
Right to Life (Art. 6)		✓
Security of Person (Art. 9)		✓
Freedom of Expression (Art. 19)	✓	✓

ICESCR

Right to Education (Art. 13)		✓
------------------------------	--	---

CEDAW

Equality in the field of education (Art. 10)		✓
--	--	---

Convention on the Rights of the Child

Right of the child to education (Arts 28-29)		✓
--	--	---

UNESCO Convention against Discrimination in Education

✓

UN Treaty Bodies: Guidance & Recommendations

Human Rights Committee

General Comment 34 on the Freedoms of Opinion and Expression (CCPR/C/GC/34)	✓	✓
General Comment No. 35 (CCPR/C/GC/35) para. 9 on Liberty and security of person		✓
General Comment No. 36 (CCPR/C/GC/36) on the Right to life		✓

Other UN Human Rights Standards

Responsibilities of Businesses

Guiding Principles on Business and Human Rights (UNGPs)	✓	✓
---	---	---

Freedom of expression, freedom of information, and the protection of journalists

Human Rights Council, Resolution 45/18 on the Safety of Journalists (A/HRC/RES/45/18)		✓
UN Special Rapporteur on freedom of opinion and expression report on hate speech online (A/74/486)		✓
UNESCO, Brisbane Declaration on Freedom of Information		✓



Decision timelines

A previous version of the Board’s Bylaws specified that the Board must make its decision on a case a maximum of 90 days from the date the case is selected for review by the Case Selection Committee. In January 2022, the Board’s Bylaws were amended so that the 90-day timeline starts when the Board announces a new case on its website.

For the two decisions the Oversight Board published in this quarter, the average time from announcement of the case to implementation of the Board’s decision has been 128 days.⁸

<i>Case ID</i>	<i>Name</i>	<i>Beginning of 90-day period</i>	<i>Board’s decision published</i>	<i>Number of days taken</i>
2022-004-FB-UA	Colombian police cartoon	5/10/22	9/15/22	128
2022-005-FB-UA	Mention of the Taliban in news reporting	5/10/22	9/15/22	128

Questions for Meta

To assist with making its decisions, the Oversight Board sends questions to Meta. Of the 60 questions sent by the Oversight Board to Meta about decisions published in this quarter, Meta answered 54 questions (90%), partially answered five questions (8%), and did not answer one question (2%).

The question that Meta did not answer in this quarter was asked as part of the “Mention of the Taliban in news reporting” case. It asked Meta to provide data on the volume of content that is removed under the Dangerous Individuals and Organizations policy through automation and the volume that it removed through human review. Meta said it was unable to verify the requested data in the time available.

⁸ The “Colombian police cartoon” case went over the 90-day deadline due to Meta requiring additional time to complete data-validation and to finalize confidentiality designations. There were also challenges scheduling Board Member deliberations, alongside other operational challenges, impacting the “Colombian police cartoon” case and the “Mention of the Taliban in news reporting” case.



Oversight Board questions answered by Meta

Number of questions

<i>Case ID</i>	<i>Name</i>	<i>Answered</i>	<i>Partially answered</i>	<i>Did not answer</i>	<i>Total</i>
2022-004-FB-UA	Colombian police cartoon	24	2	0	26
2022-005-FB-UA	Mention of the Taliban in news reporting	30	3	1	34
Total		54	5	1	60

Information around wider context of Board's decisions

In the Board's first [quarterly transparency reports](#), published in October 2021 (page 11), we explained that in the wake of disclosures around its cross-check program, Meta agreed to provide information about the wider context which may be relevant to the Board's case decisions. In that report we committed to provide further analysis in our subsequent transparency reporting on whether Meta is fulfilling this commitment.

This quarter, we have seen a welcome, continued increase in the level of technical detail Meta has provided the Board as part of their initial submissions on cases, or in response to Board questions. This includes more extensive enforcement histories, including useful explanations of mistake-prevention systems, the strikes and penalties system, and how media matching banks work. Such information provides the Board with greater insight into how Meta approaches content moderation at scale, and it has enabled the Board to issue decisions that increase transparency and make recommendations to achieve a broader impact.



Public comments

The Oversight Board conducts a public comment process to assist in its decision making.

For the two decisions published in Q3 2022, the Board received 10 comments, of which nine were published. One comment was submitted without consent to publish. The majority of public comments (60%) came from organizations, while a minority came from individuals (40%).

Public comments received by publication status

Number of comments

<i>Case ID</i>	<i>Name</i>	<i>Comments published</i>	<i>Comments not published (no consent)</i>	<i>Comments not published (violated terms)</i>	<i>Total</i>	<i>Comments unattributed⁹</i>
2022-004-FB-UA	Colombian police cartoon	3	1	0	4	0
2022-005-FB-UA	Mention of the Taliban in news reporting	6	0	0	6	3
Total		9	1	0	10	3

Public comments received by commenter type

Number of comments

<i>Case ID</i>	<i>Name</i>	<i>Individual comments</i>	<i>Organizational comments</i>	<i>Total</i>
2022-004-FB-UA	Colombian police cartoon	1	3	4
2022-005-FB-UA	Mention of the Taliban in news reporting	3	3	6
Total		4	6	10

⁹ Unattributed comments are published comments with the author's name redacted by request.



Public comments received by region

<i>Region</i>	Colombian police cartoon (2022-004-FB-UA)	Mention of the Taliban in news reporting (2022-005-FB-UA)	<i>Total</i>
United States & Canada	2	1	3
Europe	0	4	4
Asia Pacific & Oceania	0	1	1
Latin America & Caribbean	2	0	2
Sub-Saharan Africa	0	0	0
Middle East & North Africa	0	0	0
Central & South Asia	0	0	0
Total	4	6	10

40% of the public comments received for decisions published in this quarter came from Europe. A third (30%) came from the US & Canada, a fifth (20%) from Latin America and the Caribbean, and 10% from Asia Pacific and Oceania.

Recommendations

In addition to providing decisions on appealed content, the Oversight Board also made 10 policy recommendations to Meta, which the company responded to publicly within 60 days.

Of these 10 recommendations, two related to content policy (clarification of or changes to rules), seven related to enforcement (clarification of or changes to how rules are applied), and one related to transparency (on disclosure of information to the public).

Oversight Board recommendations to Meta

Number of recommendations

<i>Case ID</i>	<i>Name</i>	<i>Content policy</i>	<i>Enforcement</i>	<i>Transparency</i>	<i>Total</i>
2022-004-FB-UA	Colombian police cartoon	0	2	1	3
2022-005-FB-UA	Mention of the Taliban in news reporting	2	5	0	7
Total		2	7	1	10



Meta’s responses

Of the 10 recommendations the Board made in Q3 2022, Meta said it was implementing two “fully,” implementing two “in part,” and “assessing feasibility” on five. Meta said it would take “no further action” on one recommendation.

The Board’s assessment of Meta’s responses to our recommendations

Meta’s initial responses to the recommendations were 90% ‘comprehensive’ (nine recommendations), aside from one which was ‘somewhat comprehensive.’ For the six quarters to Q2 2022, the share of Meta’s responses falling into these two categories increased – rising from 91% in Q1 2022 to 100% in Q2 2022, and it remained at 100% in Q3 2022. This demonstrates that overall, Meta has increasingly been committing to concrete action under a set timeline, and has understood the core concern underlying the Board’s recommendations.

Board’s analysis of Meta’s implementation of Q3 recommendations

Of the 10 recommendations published in Q3 2022:

- Almost all recommendations were classified as “progress reported.” This means that Meta has committed to implementing these recommendations, but has not declared implementation to be complete, and the Board also has no evidence of implementation.
- “Colombian police cartoon” recommendation three, which asked Meta to share error rates for content mistakenly included in Media Matching Service banks of violating content, was declined.

As of Meta’s Q3 2022 update, the Board has issued 140 recommendations. Meta has 60 days to respond. The 60-day limit has not passed for 12 of the Board’s recommendations, so Meta has responded to 128 of the Board’s recommendations. It has committed to implement or implemented the majority of those recommendations.

The Board has assessed that 24 out of 128 recommendations (18.75 %) have been implemented, as demonstrated through published information. The Board assessed a further 11 recommendations (8.5%) as partially implemented, while for 43 recommendations (33.5%) Meta had reported progress towards implementation and we will continue to monitor implementation.

Implementation Category	No. of recommendations
Implementation demonstrated through published information: <i>Meta provided sufficient data for the Board to verify the recommendation has been implemented</i>	24
Partial implementation demonstrated through published information: <i>Meta has implemented a central component of the recommendation and has provided sufficient data to verify this to the Board.</i>	11
Progress reported: <i>Meta committed to implementation but has not yet completed all necessary actions.</i>	43



Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation: <i>Meta says it implemented the recommendation but has not provided sufficient evidence to verify this.</i>	28
Recommendation declined after feasibility assessment: <i>Meta engaged with the recommendation and then decided to decline its implementation after providing information on its decision.</i>	9
Recommendation omitted, declined, or reframed: <i>Meta will take no further action on the recommendation.</i>	13
Total Number of Recommendations	128

New implementation successes include:

- Meta launched new notifications globally detailing specific policy violations for the Hate Speech, Dangerous Individuals and Organizations, and Bullying & Harassment policies. It is working to expand the messaging to all Community Standards and to multiple languages by the end of 2022. This follows seven Board recommendations (“Armenians in Azerbaijan,” recommendation one, “Breast cancer symptoms and nudity,” recommendation two, “Nazi quote,” recommendation one, “Depiction of Zwarte Piet,” recommendation two, “South Africa slurs” recommendation one, “Ayahuasca brew,” recommendation two, and “Sharing private residential information” policy advisory opinion, recommendation 17). The Board considers these seven recommendations to be partially implemented. In March of this year, the Board met with Meta product managers and data scientists responsible for implementing these recommendations. In this meeting, Meta shared draft language for more granular notifications with the Board (see Annex III). The Board will consider the recommendation fully implemented when Meta confirms the notifications have been rolled out in all languages used on Meta’s platforms, and on Instagram.
- Due to a series of recommendations from the Board referencing the Dangerous Organizations and Individuals policy, Meta initiated an in-depth Dangerous Organizations and Individuals policy review focused on how it can prioritize designations of organizations and individuals within this policy based on risk. It is also engaging in a policy development process on its framework for assessing whether content amounts to praise, substantive support, or representation of a designated individual or organization. This demonstrates the Board’s broader systemic impact that goes beyond recommendations on Meta’s approach to the Dangerous Individuals and Organizations policy area.
- Following the Board’s recommendation in the “Ayahuasca brew” case, Meta has added definitions of the terms “non-medical drugs” and “pharmaceutical drugs” to its Restricted Goods and Services policy.
- Following the Board’s recommendations made in its decision in a case about a post describing sexual assault against minors (2021-016-FB-FBR), Meta has initiated an in-depth policy review of its approach to preventing the identification of victims of sexual violence in its Child Sexual Exploitation, Abuse and Nudity and Adult Sexual Exploitation policies.



- Meta also acknowledged that it had previously not addressed “Breast cancer symptoms and nudity,” recommendation seven (that Meta revise the “short” explanation of the Instagram Community Guidelines to clarify that the ban on adult nudity is not absolute), and is now tracking it along with recommendation nine from the same case. Because Meta originally failed to respond to this recommendation, it was assessed as “recommendation omitted, declined, or reframed.” After this response, we have reassessed the recommendation as “progress reported,” consistent with the “Breast cancer symptoms and nudity” recommendation nine, which will be completed when Meta brings Facebook and Instagram’s policies into alignment.
- Meta’s original response to “Asking for Adderall,” recommendation three reframed the recommendation in such a way that Meta could say it already discloses this type of information in its Community Standards Enforcement Report. Meta also originally rejected disclosure of accuracy metrics as a meaningful transparency initiative, stating that they do not “provide a complete picture of our content moderation successes and challenges.” After this initial response, the recommendation was assessed to be, “omitted, declined, or reframed”. However, Meta’s Q3 update to this recommendation now indicates that it is working to define reviewer accuracy metrics – the Board now considers this recommendation to be “in progress.”
- In the “Former President Trump’s suspension” case, the Board initially tracked no response from Meta for recommendation 12. However, Meta recently informed the Board that their response was bundled with their response to another recommendation (“Former President Trump’s suspension,” recommendation 13). After re-reading this response, the Board agrees that it addressed the recommendation previously thought to be omitted. Therefore, the Board is now assessing recommendation 12 alongside recommendation 13, which it assesses as: “Meta has reported implementation or described as work Meta already does, but did not publish information to demonstrate implementation.”



Annex I - Summaries of cases where Meta identified its original decision on a piece of content was incorrect

The Board's shortlist of cases is sent to Meta's legal team to review for eligibility. In practice, at this stage, Meta has also assessed whether its original decision on a piece of content was correct or not. In Q3 2022, Meta determined that its original decision on the piece of content was incorrect in the following 12 of the 20 cases shortlisted by the Board.

1. Even though Meta acknowledged its original decision in this case as incorrect, the case was selected by the Oversight Board as the case "Iran protest slogan" case. Read a description of the post [here](#).
2. The content contains text and a picture quoting from the Qur'an, referring to "disbelievers." It was taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it seems that the removal was in error because the content contains no discernible hate speech.
3. The content contains a quote from, and a picture of, Amilcar Cabral. In the quote, Cabral states that it will be demonstrated that colonialists took colonized peoples out of the latter's own history to follow the history of the former. It was taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it is unclear if Meta treats Cabral as a Dangerous Individual, and whether the content constitutes praise, representation or substantive support of a Dangerous Individual.
4. The content is from the verified account of a Palestinian journalist. His post is addressing the ongoing escalations in Gaza, specifically the bombings that resulted in the death of Tayseer Al-Jabari, a prominent leader of the military arm of the Islamic Jihad Movement. In his post he claims not to be an advocate for war, but for his audience not to blame the Islamic Jihad Movement for the violence. He ends his post by stating that he hopes that the aggression ends soon, and that peace is achieved. In his post he mentions the victim of the violence, Tayseer Al-Jabari, by name. It was taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it is unclear if Meta treats Al-Jabari as a Dangerous Individual, and whether the content constitutes praise, representation or substantive support of a Dangerous Individual or Organization, such as through the call "not to blame" the Islamic Jihad Movement.
5. The content contains a video showing a man holding up a "Black Lives Matter" sign in Harrison, Arkansas (the home of the Ku Klux Klan headquarters) and recording the reactions of people driving by. Most of the reactions include explicit language and racial slurs. In their statement, the user



explained that this was a piece of undercover journalism about racism in the United States and the video did not in any way advocate for the behavior or words of those depicted. It was taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it seems that the content does not constitute praise, representation or substantive support of a Dangerous Individual or Organization.

6. The content contains text that describes in a sarcastic way all the ways in which women would be upset if men stopped harassing them. The user writes that men think that women will be depressed when they stop receiving attention from men, but in actuality, women will stand to benefit in many ways. It was taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it seems that the removal was in error because the content contains no discernible hate speech.
7. The content was posted before the 2022 Italian general election. It describes certain rapes happening in the United Kingdom and claims that Muslims and Pakistanis were the perpetrators. The text also uses slurs when referring to abusers and victims. It links impunity for these crimes to the political left. The content was kept up. The Case Selection Committee referred this appeal to Meta, who removed the content. Without explanation from Meta on either the keep-up or the removal, it seems that the content does contain hate speech at least because of the slurs used.
8. The content is a video made by Al Jazeera regarding an attack in the Democratic Republic of the Congo and uploaded by an individual Facebook user. The video shows the bodies of victims, including infants and children, in a violent attack. The content was taken down under the Community Standard on Child Sexual Exploitation, Abuse and Nudity. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it seems that the removal was in error because the private parts on the infants' and children's bodies are blurred in the video.
9. The content criticizes what it calls the 'man-world' and contains an image of an alleged victim of rape and extra-judicial execution perpetrated by Sri Lanka government security forces. The content also includes a picture of a young female politician who allegedly went through a similar experience. The content was taken down under the Community Standard on Adult Sexual Exploitation. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it seems that nothing in the content falls under Adult Sexual Exploitation.
10. The content is a post from Brazil which describes the various ways in which men sexually abuse and exploit women, such as mistreating their partners, taking drunk girls home, taking explicit photos of women without their consent and sharing them to WhatsApp groups. The post notes that rape and abuse are huge problems in Brazil, describing the history of rape against Indigenous and enslaved women by Europeans. The post also addresses the "hypocritical man" and states that most of his



friends are rapists. The content was taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it is unclear whether the statement that most men are rapists falls under an exception of Tier 1 of the Community Standard on Hate Speech, which allows for critical speech targeted at “non-protected groups described as having carried out violent crimes or sexual offenses....”

11. The content is a post from Ukraine describing the various forms of sexual harassment that women encounter on a daily basis throughout the course of their lifetimes; from having men's genitals brush up against them on public transit, to being the victims of incest from family members and sexual abuse from others in their communities such as teachers and sports coaches. The post describes men as being pedophiles and rapists, and that they are the reason women are lured into prostitution and pornography. The post ends with the user lamenting how the war in Ukraine has shown that women are not insulated from becoming sex slaves or falling into prostitution. The content was taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Meta, who restored the content. Without explanation from Meta on either the removal or the restoration, it is unclear whether the reference to men as pedophiles and rapists falls into an exception under Tier 1 of the Community Standard on Hate Speech, as it might be treated as targeting a group of people “who are considered non-protected groups described as having carried out violent crimes or sexual offences or representing less than half of a group”.
12. The content is a post from Thailand criticizing the Thai King and a judge who prohibited a young woman from traveling abroad. The post claims that courts in Thailand are prohibiting activists from leaving the country for overseas studies. The content was taken down under the Community Standard on Bullying and Harassment. The Case Selection Committee referred this appeal to Meta, who restored the content but withheld the appeal from the Board for legal reasons, citing legal obligations in Thailand.



Annex II – The Board’s assessment of how Meta has responded to and implemented its recommendations

Recommendation identifier	Quarter	Recommendation	Category	Board's Assessment of Meta's response	Board's assessment of implementation
Armenians in Azerbaijan #1	Q12021	Ensure that users are always notified of the reasons for any enforcement of the Community Standards against them, including the specific rule Facebook is enforcing. Doing so would enable Facebook to encourage expression that complies with its Community Standards, rather than adopting an adversarial posture towards users. In this case, the user was informed that the post violated the Community Standard on hate speech but was not told that the post violated the standard because it included a slur targeting national origin. Facebook satisfied the principle of legality in this instance, but Facebook's lack of transparency left its decision susceptible to the mistaken belief that it had removed the post because the user was addressing a controversial subject or expressing a viewpoint Facebook disagreed with.	Enforcement	Somewhat Comprehensive	Partial Implementation demonstrated through published information
Breast cancer symptoms and nudity #1	Q12021	Improve the automated detection of images with text-overlay to ensure that posts raising awareness of breast cancer symptoms are not wrongly flagged for review.	Enforcement	Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to



Breast cancer symptoms and nudity #2	Q12021	Ensure that users are always notified of the reasons for the enforcement of content policies against them, providing the specific rule within the Community Standard Facebook based its decision on.	Enforcement	Not Comprehensive	demonstrate implementation Partial Implementation demonstrated through published information
Breast cancer symptoms and nudity #3	Q12021	Inform users when automation is used to take enforcement action against their content, including accessible descriptions of what this means.	Enforcement	Somewhat Comprehensive	Progress reported
Breast cancer symptoms and nudity #4	Q12021	Ensure users can appeal decisions taken by automated systems to human review when their content is found to have violated Facebook's Community Standard on Adult Nudity and Sexual Activity. Where Facebook is seeking to prevent child sexual exploitation or the dissemination of non-consensual intimate images, it should enforce based on its Community Standards on Sexual Exploitation of Adults and Child Sexual Exploitation, Abuse and Nudity, rather than rely on over-enforcing policies on adult nudity. Appeals should still be available in these cases, so incorrect removals of permitted consensual adult nudity can be reversed.	Enforcement	Not Comprehensive	Recommendation declined after feasibility assessment



Breast cancer symptoms and nudity #5	Q12021	Implement an internal audit procedure to continuously analyze a statistically representative sample of automated content removal decisions to reverse and learn from enforcement mistakes.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
Breast cancer symptoms and nudity #6	Q12021	Expand transparency reporting to disclose data on the number of automated removal decisions per Community Standard, and the proportion of those decisions subsequently reversed following human review.	Transparency	Somewhat Comprehensive	Progress reported
Breast cancer symptoms and nudity #7	Q12021	Revise the “short” explanation of the Instagram Community Guidelines to clarify that the ban on adult nudity is not absolute.	Content Policy	Not Comprehensive	Progress reported
Breast cancer symptoms and nudity #8	Q12021	Revise the “long” explanation of the Instagram Community Guidelines to clarify that visible female nipples can be shown to raise breast cancer awareness.	Content Policy	Comprehensive	Implementation demonstrated through published information
Breast cancer symptoms and nudity #9	Q12021	Clarify that the Instagram Community Guidelines are interpreted in line with the Facebook Community Standards, and where there are inconsistencies the latter take precedence.	Content Policy	Comprehensive	Progress reported
Nazi Quote #1	Q12021	Ensure that users are always notified of the reasons for any enforcement of the Community Standards against them, including the specific rule Facebook is enforcing (e.g. for support of a hate organization).	Enforcement	Somewhat Comprehensive	Partial Implementation demonstrated through published information



<p>Nazi Quote #2</p> <p>Q12021</p>	<p>Explain and provide examples of the application of key terms used in the Dangerous Individuals and Organizations policy, including the meanings of “praise,” “support” and “representation.” These should align with the definitions used in Facebook’s Internal Implementation Standards. The Community Standard should provide clearer guidance to users on how to make their intent apparent when discussing individuals or organizations designated as dangerous.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Nazi Quote #3</p> <p>Q12021</p>	<p>Provide a public list of the organizations and individuals designated “dangerous” under the Dangerous Individuals and Organizations Community Standard. At a minimum, illustrative examples should be provided. This would help users to better understand the policy and conduct themselves accordingly.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>Claimed COVID-19 cure #1</p> <p>Q12021</p>	<p>The Board recommends that Facebook set out a clear and accessible Community Standard on health misinformation, consolidating and clarifying existing rules in one place (including defining key terms such as misinformation). This rule-making should be accompanied with “detailed hypotheticals that illustrate the nuances of interpretation and application of [these] rules” to provide further clarity for users (See report A/HRC/38/35, para. 46 (2018)). Facebook should conduct a human rights impact assessment with relevant stakeholders as part of its process of rule modification (UNGPs, Principles 18-19).</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



Claimed COVID-19 cure #2

Q12021

To ensure enforcement measures on health misinformation represent the least intrusive means of protecting public health, the Board recommends that Facebook: Clarify the particular harms it is seeking to prevent and provide transparency about how it will assess the potential harm of particular content; Conduct an assessment of its existing range of tools to deal with health misinformation; Consider the potential for development of further tools that are less intrusive than content removals; Publish its range of enforcement options within the Community Standards, ranking these options from most to least intrusive based on how they infringe freedom of expression; Explain what factors, including evidence-based criteria, the platform will use in selecting the least intrusive option when enforcing its Community Standards to protect public health; Make clear within the Community Standards what enforcement option applies to each rule

Content Policy

Comprehensive

Implementation demonstrated through published information



Claimed COVID-19 cure #3

Q12021

In cases where users post information about COVID-19 treatments that contradicts the specific advice of health authorities and where a potential for physical harm is identified but is not imminent, the Board strongly recommends Facebook to adopt a range of less intrusive measures. This could include labelling which alerts users to the disputed nature of the post's content and provides links to the views of the World Health Organization and national health authorities. In certain situations it may be necessary to introduce additional friction to a post - for example, by preventing interactions or sharing, to reduce organic and algorithmically driven amplification. Downranking content, to prevent visibility in other users' newsfeeds, might also be considered. All enforcement measures, including labelling or other methods of introducing friction, should be clearly communicated to users, and subject to appeal.

Enforcement

Not Comprehensive

Recommendation omitted, declined, or reframed



<p>Claimed COVID-19 cure #4</p> <p>Q12021</p>	<p>Publish a transparency report on how the Community Standards have been enforced during the COVID-19 global health crisis. This should include: data in absolute and percentage terms on the number of removals, as well as data on other enforcement measures, on the specific Community Standards enforced against, including on the proportion that relied entirely on automation; a breakdown by content type enforced against (including individual posts, accounts, and groups); a breakdown by the source of detection (including automation, user flagging, trusted partners, law enforcement authorities); a breakdown by region and language; metrics on the effectiveness of less intrusive measures (e.g., impact of labelling or downranking); data on the availability of appeals throughout the crisis, including the total number of cases where appeal was withdrawn entirely, and the percentage of automated appeals; conclusions and lessons learned, including information on any changes Facebook is making to ensure greater compliance with its human rights responsibilities going forward.</p>	<p>Transparency</p>	<p>Not Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Protest in India against France #1</p> <p>Q12021</p>	<p>To ensure users have clarity regarding permissible content, the Board recommends that Facebook provide users with additional information regarding the scope and enforcement of this Community Standard. Enforcement criteria should be public and align with Facebook’s Internal Implementation Standards. Specifically, Facebook’s criteria should address intent, the identity of the user and audience, and context.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



<p>Depiction of Zwarte Piet #1</p> <p>Q22021</p>	<p>Facebook should link the rule in the Hate Speech Community Standard prohibiting blackface to the company’s reasoning for the rule, including harms it seeks to prevent.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Depiction of Zwarte Piet #2</p> <p>Q22021</p>	<p>In line with the Board’s recommendation in case 2020-003-FB-UA, Facebook should “ensure that users are always notified of the reasons for any enforcement of the Community Standards against them, including the specific rule Facebook is enforcing.” In this case any notice to users should specify the rule on blackface, and also link to above mentioned resources that explain the harm this rule seeks to prevent. Facebook should provide a detailed update on its “feasibility assessment” of the Board’s prior recommendations on this topic, including the specific nature of any technical limitations and how these can be overcome.</p>	<p>Enforcement</p>	<p>Somewhat Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>
<p>Punjabi concern over the RSS in India #1</p> <p>Q22021</p>	<p>Facebook should translate its Community Standards and Internal Implementation Standards into Punjabi. Facebook should aim to make its Community Standards accessible in all languages widely spoken by its users. This would allow a full understanding of the rules that users must abide by when using Facebook’s products. It would also make it simpler for users to engage with Facebook over content that may violate their rights.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



<p>Punjabi concern over the RSS in India #2</p>	<p>Q22021</p>	<p>In line with the Board’s recommendation in case 2020-004-IG-UA, the company should restore human review and access to a human appeals process to pre-pandemic levels as soon as possible while fully protecting the health of Facebook’s staff and contractors.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Punjabi concern over the RSS in India #3</p>	<p>Q22021</p>	<p>Facebook should improve its transparency reporting to increase public information on error rates by making this information viewable by country and language for each Community Standard. The Board underscores that more detailed transparency reports will help the public spot areas where errors are more common, including potential specific impacts on minority groups, and alert Facebook to correct them.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Progress reported</p>



Former President
Trump's
suspension #1

Q22021

The Board believes that it is not always useful to draw a firm distinction between political leaders and other influential users. It is important to recognise that other users with large audiences can also contribute to serious risks of harm. The same rules should apply to all users of the platform; but context matters when assessing issues of causality and the probability and imminence of harm. What is important is the degree of influence that a user has over other users [...] Facebook must assess posts by influential users in context according to the way they are likely to be understood, even if their incendiary message is couched in language designed to avoid responsibility, such as superficial encouragement to act peacefully or lawfully. Facebook used the six contextual factors in the Rabat Plan of Action in this case and the Board thinks that this is a useful way to assess the contextual risks of potentially harmful speech. The Board stresses that time is of the essence in such situations; taking action before influential users can cause significant harm should take priority over newsworthiness and other values of political communication.

Content
Policy

Somewhat
Comprehensive

Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



<p>Former President Trump's suspension #2</p> <p>Q22021</p>	<p>When posts by influential users pose a high probability of imminent harm, as assessed under international human rights standards, Facebook should take action to enforce its rules quickly.</p>	<p>Enforcement</p>	<p>Somewhat Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Former President Trump's suspension #3</p> <p>Q22021</p>	<p>While all users should be held to the same content policies, there are unique factors that must be considered in assessing the speech of political leaders. Heads of state and other high-ranking government officials can have a greater power to cause harm than other people. Facebook should recognize that posts by heads of state and other high officials of government can carry a heightened risk of encouraging, legitimizing, or inciting violence - either because their high position of trust imbues their words with greater force and credibility or because their followers may infer they can act with impunity. At the same time, it is important to protect the rights of people to hear political speech. Nonetheless, if the head of state or high government official has repeatedly posted messages that pose a risk of harm under international human rights norms, Facebook should suspend the account for a determinate period sufficient to protect against imminent harm.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>



Former President Trump's suspension #4	Q22021	Periods of suspension should be long enough to deter misconduct and may, in appropriate cases, include account or page deletion.	Content Policy	Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
Former President Trump's suspension #5	Q22021	Restrictions on speech are often imposed by or at the behest of powerful state actors against dissenting voices and members of political oppositions. Facebook must resist pressure from governments to silence their political opposition. When assessing potential risks, Facebook should be particularly careful to consider the relevant political context.	Enforcement	Somewhat Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
Former President Trump's suspension #6	Q22021	In evaluating political speech from highly influential users, Facebook should rapidly escalate the content moderation process to specialized staff who are familiar with the linguistic and political context and insulated from political and economic interference and undue influence. This analysis should examine the conduct of highly influential users off the Facebook and Instagram platforms to adequately assess the full relevant context of potentially harmful speech. Further, Facebook should ensure that it dedicates adequate resourcing and expertise to assess risks of harm from influential accounts globally.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



Former President Trump's suspension #7	Q22021	Facebook should publicly explain the rules that it uses when it imposes account-level sanctions against influential users. These rules should ensure that when Facebook imposes a time-limited suspension on the account of an influential user to reduce the risk of significant harm, it will assess whether the risk has receded before the suspension term expires. If Facebook identifies that the user poses a serious risk of inciting imminent violence, discrimination, or other lawless action at that time, another time-bound suspension should be imposed when such measures are necessary to protect public safety and proportionate to the risk.	Content Policy	Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
Former President Trump's suspension #8	Q22021	When Facebook implements special procedures that apply to influential users, these should be well documented. It was unclear whether Facebook applied different standards in this case, and the Board heard many concerns about the potential application of the newsworthiness allowance. It is important that Facebook address this lack of transparency and the confusion it has caused.	Transparency	Comprehensive	Implementation demonstrated through published information
Former President Trump's suspension #9	Q22021	Facebook should produce more information to help users understand and evaluate the process and criteria for applying the newsworthiness allowance. Facebook should clearly explain how the newsworthiness allowance applies to influential accounts, including political leaders and other public figures.	Content Policy	Comprehensive	Implementation demonstrated through published information



Former President Trump's suspension #10	Q22021	For cross check review, Facebook should clearly explain the rationale, standards, and processes of review, including the criteria to determine which pages and accounts are selected for inclusion.	Content Policy	Somewhat Comprehensive	Implementation demonstrated through published information
Former President Trump's suspension #11	Q22021	Facebook should report on the relative error rates and thematic consistency of determinations made through the cross-check process compared with ordinary enforcement procedures.	Transparency	Not Comprehensive	Recommendation omitted, declined, or reframed
Former President Trump's suspension #12	Q22021	When Facebook's platform has been abused by influential users in a way that results in serious adverse human rights impacts, it should conduct a thorough investigation into the incident. Facebook should assess what influence it had and assess what changes it could enact to identify, prevent, mitigate, and account for adverse impacts in future.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
Former President Trump's suspension #13	Q22021	Facebook should undertake a comprehensive review of its potential contribution to the narrative of electoral fraud and the exacerbated tensions that culminated in the violence in the United States on January 6, 2021. This should be an open reflection on the design and policy choices that Facebook has made that may enable its platform to be abused. Facebook should carry out this due diligence, implement a plan to act upon its findings, and communicate openly about how it addresses adverse human rights impacts it was involved with.	Transparency	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



<p>Former President Trump's suspension #14</p> <p>Q22021</p>	<p>Facebook has a responsibility to collect, preserve and, where appropriate, share information to assist in the investigation and potential prosecution of grave violations of international criminal, human rights and humanitarian law by competent authorities and accountability mechanisms. Facebook's corporate human rights policy should make clear the protocols the company has in place in this regard. The policy should also make clear how information previously public on the platform can be made available to researchers conducting investigations that conform with international standards and applicable data protection laws.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Former President Trump's suspension #15</p> <p>Q22021</p>	<p>Facebook should explain in its Community Standards and Guidelines its strikes and penalties process for restricting profiles, pages, groups and accounts on Facebook and Instagram in a clear, comprehensive, and accessible manner. These policies should provide users with sufficient information to understand when strikes are imposed (including any applicable exceptions or allowances) and how penalties are calculated.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Former President Trump's suspension #16</p> <p>Q22021</p>	<p>Facebook should also provide users with accessible information on how many violations, strikes, and penalties have been assessed against them, as well as the consequences that will follow future violations.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



<p>Former President Trump's suspension #17</p> <p>Q22021</p>	<p>In its transparency reporting, Facebook should include numbers of profile, page, and account restrictions, including the reason and manner in which enforcement action was taken, with information broken down by region and country.</p>	<p>Transparency</p>	<p>Not Comprehensive</p>	<p>Progress reported</p>
<p>Former President Trump's suspension #18</p> <p>Q22021</p>	<p>Facebook should develop and publish a policy that governs its response to crises or novel situations where its regular processes would not prevent or avoid imminent harm. While these situations cannot always be anticipated, Facebook's guidance should set appropriate parameters for such actions, including a requirement to review its decision within a fixed time.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>
<p>"Two buttons" meme #1</p> <p>Q22021</p>	<p>Make technical arrangements to ensure that notice to users refers to the Community Standard enforced by the company. If Facebook determines that (i) the content does not violate the Community Standard notified to user, and (ii) that the content violates a different Community Standard, the user should be properly notified about it and given another opportunity to appeal. They should always have access to the correct information before coming to the Board.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>"Two buttons" meme #2</p> <p>Q22021</p>	<p>Include the satire exception, which is currently not communicated to users, in the public language of the Hate Speech Community Standard.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



"Two buttons" meme #3	Q22021	Make sure that it has adequate procedures in place to assess satirical content and relevant context properly. This includes providing content moderators with: (i) access to Facebook's local operation teams to gather relevant cultural and background information; and (ii) sufficient time to consult with Facebook's local operation teams and to make the assessment. Facebook should ensure that its policies for content moderators incentivize further investigation or escalation where a content moderator is not sure if a meme is satirical or not.	Enforcement	Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
"Two buttons" meme #4	Q22021	Let users indicate in their appeal that their content falls into one of the exceptions to the Hate Speech policy. This includes exceptions for satirical content and where users share hateful content to condemn it or raise awareness.	Enforcement	Comprehensive	Progress reported
"Two buttons" meme #5	Q22021	Ensure appeals based on policy exceptions are prioritized for human review.	Enforcement	Comprehensive	Progress reported
Pro-Navalny protests in Russia #1	Q22021	Facebook should amend and redraft the Bullying & Harassment Community Standard to explain the relationship between the Policy Rationale and the "Do not's" as well as the other rules restricting content that follow it.	Content Policy	Comprehensive	Implementation demonstrated through published information



<p>Pro-Navalny protests in Russia #2</p> <p>Q22021</p>	<p>Differentiate between bullying and harassment and provide definitions that distinguish the two acts. Further, the Community Standard should clearly explain to users how bullying and harassment differ from speech that only causes offense and may be protected under international human rights law.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>Pro-Navalny protests in Russia #3</p> <p>Q22021</p>	<p>Clearly define its approach to different target user categories and provide illustrative examples of each target category (i.e. who qualifies as a public figure). Format the Community Standard on Bullying and Harassment by user categories currently listed in the policy.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Pro-Navalny protests in Russia #4</p> <p>Q22021</p>	<p>Include illustrative examples of violating and non-violating content in the Bullying and Harassment Community Standard to clarify the policy lines drawn and how these distinctions can rest on the identity status of the target.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>Pro-Navalny protests in Russia #5</p> <p>Q22021</p>	<p>When assessing content including a 'negative character claim' against a private adult, Facebook should amend the Community Standard to require an assessment of the social and political context of the content. Facebook should reconsider the enforcement of this rule in political or public debates where the removal of the content would stifle debate.</p>	<p>Content Policy</p>	<p>Not Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>



<p>Pro-Navalny protests in Russia #6</p> <p>Q22021</p>	<p>Whenever Facebook removes content because of a negative character claim that is only a single word or phrase in a larger post, it should promptly notify the user of that fact, so that the user can repost the material without the negative character claim.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Öcalan's isolation #1</p> <p>Q32021</p>	<p>Immediately restore the misplaced 2017 guidance to the Internal Implementation Standards and Known Questions (the internal guidance for content moderators), informing all content moderators that it exists and arranging immediate training on it.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Öcalan's isolation #2</p> <p>Q32021</p>	<p>Evaluate automated moderation processes for enforcement of the Dangerous Individuals and Organizations policy. Where necessary, Facebook should update classifiers to exclude training data from prior enforcement errors that resulted from failures to apply the 2017 guidance. New training data should be added that reflects the restoration of this guidance.</p>	<p>Enforcement</p>	<p>Not Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Öcalan's isolation #3</p> <p>Q32021</p>	<p>Publish the results of the ongoing review process to determine if any other policies were lost, including descriptions of all lost policies, the period the policies were lost for, and steps taken to restore them.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>



Öcalan's isolation #4 Q32021

Reflect in the Dangerous Individuals and Organizations "policy rationale" that respect for human rights and freedom of expression, in particular open discussion about human rights violations and abuses that relate to terrorism and efforts to counter terrorism, can advance the value of "Safety," and that it is important for the platform to provide a space for these discussions. While "Safety" and "Voice" may sometimes be in tension, the policy rationale should specify in greater detail the "real-world harms" the policy seeks to prevent and disrupt when "Voice" is suppressed.

Content Policy

Comprehensive

Implementation demonstrated through published information



Öcalan's isolation #5 Q32021

Add to the Dangerous Individuals and Organizations policy a clear explanation of what “support” excludes. Users should be free to discuss alleged violations and abuses of the human rights of members of designated organizations. This should not be limited to detained individuals. It should include discussion of rights protected by the UN human rights conventions as cited in Facebook’s Corporate Human Rights Policy. This should allow, for example, discussions on allegations of torture or cruel, inhuman, or degrading treatment or punishment, violations of the right to a fair trial, as well as extrajudicial, summary, or arbitrary executions, enforced disappearance, extraordinary rendition and revocation of citizenship rendering a person stateless. Calls for accountability for human rights violations and abuses should also be protected. Content that incites acts of violence or recruits people to join or otherwise provide material support to Facebook-designated organizations should be excluded from protection even if the same content also discusses human rights concerns. The user’s intent, the broader context in which they post, and how other users understand their post, is key to determining the likelihood of real-world harm that may result from such posts.

Content Policy

Comprehensive

Implementation demonstrated through published information



<p>Öcalan's isolation #6</p> <p>Q32021</p>	<p>Explain in the Community Standards how users can make the intent behind their posts clear to Facebook. This would be assisted by implementing the Board's existing recommendation to publicly disclose the company's list of designated individuals and organizations (see: case 2020-005-FB-UA). Facebook should also provide illustrative examples to demonstrate the line between permitted and prohibited content, including in relation to the application of the rule clarifying what "support" excludes.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>
<p>Öcalan's isolation #7</p> <p>Q32021</p>	<p>Ensure meaningful stakeholder engagement on the proposed policy change through Facebook's Product Policy Forum, including through a public call for inputs. Facebook should conduct this engagement in multiple languages across regions, ensuring the effective participation of individuals most impacted by the harms this policy seeks to prevent. This engagement should also include human rights, civil society, and academic organizations with expert knowledge on those harms, as well as the harms that may result from over-enforcement of the existing policy.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Öcalan's isolation #8</p> <p>Q32021</p>	<p>Ensure internal guidance and training is provided to content moderators on any new policy. Content moderators should be provided adequate resources to be able to understand the new policy, and adequate time to make decisions when enforcing the policy.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>



<p>Öcalan's isolation # 9</p> <p>Q32021</p>	<p>Ensure that users are notified when their content is removed. The notification should note whether the removal is due to a government request or due to a violation of the Community Standards or due to a government claiming a national law is violated (and the jurisdictional reach of any removal).</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Öcalan's isolation #10</p> <p>Q32021</p>	<p>Clarify to Instagram users that Facebook's Community Standards apply to Instagram in the same way they apply to Facebook, in line with the recommendation in case 2020-004-IG-UA.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Öcalan's isolation #11</p> <p>Q32021</p>	<p>Include information in its transparency reporting on the number of requests Facebook receives for content removals from governments that are based on Community Standards violations (as opposed to violations of national law), and the outcome of those requests.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Progress reported</p>
<p>Öcalan's isolation #12</p> <p>Q32021</p>	<p>In transparency reporting, include more comprehensive information on error rates for enforcing rules on "praise" and "support" of dangerous individuals and organizations, broken down by region and language.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>Myanmar bot #1</p> <p>Q32021</p>	<p>Facebook should ensure that its Internal Implementation Standards are available in the language in which content moderators review content. If necessary to prioritize, Facebook should focus first on contexts where the risks to human rights are more severe.</p>	<p>Enforcement</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>



COVID-19 lockdowns in Brazil #1	Q32021	Facebook should conduct a proportionality analysis to identify a range of less intrusive measures than removing the content. When necessary, the least intrusive measures should be used where content related to COVID-19 distorts the advice of international health authorities and where a potential for physical harm is identified but is not imminent. Recommended measures include: (a) labeling content to alert users to the disputed nature of a post's content and to provide links to the views of the World Health Organization and national health authorities; (b) introducing friction to posts to prevent interactions or sharing; and (c) down-ranking, to reduce visibility in other users' News Feeds. All these enforcement measures should be clearly communicated to all users, and subject to appeal.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
COVID-19 lockdowns in Brazil #2	Q32021	Given the context of the COVID-19 pandemic, Facebook should make technical arrangements to prioritize fact-checking of potential health misinformation shared by public authorities which comes to the company's attention, taking into consideration the local context.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation
COVID-19 lockdowns in Brazil #3	Q32021	Facebook should provide more transparency within the False News Community Standard regarding when content is eligible for fact-checking, including whether public institutions' accounts are subject to fact-checking.	Content Policy	Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



<p>Shared Al Jazeera post #1</p> <p>Q32021</p>	<p>Add criteria and illustrative examples to its Dangerous Individuals and Organizations policy to increase understanding of the exceptions for neutral discussion, condemnation and news reporting.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Shared Al Jazeera post #2</p> <p>Q32021</p>	<p>Ensure swift translation of updates to the Community Standards into all available languages.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation declined after feasibility assessment</p>
<p>Shared Al Jazeera post #3</p> <p>Q32021</p>	<p>Engage an independent entity not associated with either side of the Israeli-Palestinian conflict to conduct a thorough examination to determine whether Facebook’s content moderation in Arabic and Hebrew, including its use of automation, have been applied without bias. This examination should review not only the treatment of Palestinian or pro-Palestinian content, but also content that incites violence against any potential targets, no matter their nationality, ethnicity, religion or belief, or political opinion. The review should look at content posted by Facebook users located in and outside of Israel and the Palestinian Occupied Territories. The report and its conclusions should be made public.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>



Shared Al Jazeera post #4	Q32021	Formalize a transparent process on how it receives and responds to all government requests for content removal, and ensure that they are included in transparency reporting. The transparency reporting should distinguish government requests that led to removals for violations of the Community Standards from requests that led to removal or geo-blocking for violating local law, in addition to requests that led to no action.	Transparency	Comprehensive	Progress reported
Colombia protests #1	Q32021	Publish illustrative examples from the list of slurs designated as violating under its Hate Speech Community Standard, including borderline cases with words which may be harmful in some contexts but not others.	Content Policy	Comprehensive	Recommendation declined after feasibility assessment
Colombia protests #2	Q32021	Link the short explanation of the newsworthiness allowance provided in the introduction to the Community Standards to the more detailed Transparency Center explanation of how this policy applies. The company should supplement this explanation with illustrative examples from a variety of contexts, including reporting on large scale protests.	Content Policy	Comprehensive	Implementation demonstrated through published information
Colombia protests #3	Q32021	Develop and publicize clear criteria for content reviewers for escalating for additional review public interest content that potentially violates the Community Standards but may be eligible for the newsworthiness allowance. These criteria should cover content depicting large protests on political issues.	Enforcement	Not Comprehensive	Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



<p>Colombia protests #4</p> <p>Q32021</p>	<p>Notify all users who reported content which was assessed as violating but left on the platform for public interest reasons that the newsworthiness allowance was applied to the post. The notice should link to the Transparency Center explanation of the newsworthiness allowance.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>South Africa slurs #1</p> <p>Q32021</p>	<p>Notify users of the specific rule within the Hate Speech Community Standard that has been violated in the language in which they use Facebook, as recommended in case decision 2020-003-FB-UA (Armenians in Azerbaijan) and case decision 2021-002-FB-UA (Depiction of Zwarte Piet). In this case, for example, the user should have been notified they violated the slurs prohibition. The Board has noted Facebook’s response to Recommendation No. 2 in case decision 2021-002-FB-UA, which describes a new classifier that should be able to notify English-language Facebook users their content has violated the slur rule. The Board looks forward to Facebook providing information that confirms implementation for English-language users and information about the timeframe for implementation for other language users.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>



<p>Wampum belt #1 Q42021</p>	<p>Provide users with timely and accurate notice of any company action being taken on the content their appeal relates to. Where applicable, including in enforcement error cases like this one, the notice to the user should acknowledge that the action was a result of the Oversight Board’s review process. Meta should share the user messaging sent when Board actions impact content decisions appealed by users, to demonstrate it has complied with this recommendation. These actions should be taken with respect to all cases that are corrected at the eligibility stage of the Board’s process.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Wampum belt #2 Q42021</p>	<p>Study the impacts of modified approaches to secondary review on reviewer accuracy and throughput. In particular, the Board requests an evaluation of accuracy rates when content moderators are informed that they are engaged in secondary review, so they know the initial determination was contested. This experiment should ideally include an opportunity for users to provide relevant context that may help reviewers evaluate their content, in line with the Board’s previous recommendations. Meta should share the results of these accuracy assessments with the Board and summarize the results in its quarterly Board transparency report to demonstrate it has complied with this recommendation.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>



Wampum belt #3 Q42021

Conduct accuracy assessments focused on Hate Speech policy allowances that cover artistic expression and expression about human rights violations (e.g., condemnation, awareness raising, self-referential use, empowering use). This assessment should also specifically investigate how the location of a reviewer impacts the ability of moderators to accurately assess hate speech and counter speech from the same or different regions. The Board understands this analysis likely requires the development of appropriate and accurately labelled samples of relevant content. Meta should share the results of this assessment with the Board, including how these results will inform improvements to enforcement operations and policy development and whether it plans to run regular reviewer accuracy assessments on these allowances, and summarize the results in its quarterly Board transparency report to demonstrate it has complied with this recommendation.

Enforcement

Somewhat
Comprehensive

Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation



<p>Ayahuasca brew #1</p> <p>Q42021</p>	<p>The Board reiterates its recommendation from case decision 2020-004-IG-UA and case decision 2021-006-IG-UA that Meta should explain to users that it enforces the Facebook Community Standards on Instagram, with several specific exceptions. The Board notes Meta’s response to these recommendations. While Meta may be taking other actions to comply with the recommendations, the Board recommends Meta update the introduction to the Instagram Community Guidelines (“The Short” Community Guidelines) within 90 days to inform users that if content is considered violating on Facebook, it is also considered violating on Instagram, as stated in the company’s Transparency Center, with some exceptions.</p>	<p>Enforcement</p>	<p>Somewhat Comprehensive</p>	<p>Progress reported</p>
<p>Ayahuasca brew #2</p> <p>Q42021</p>	<p>The Board reiterates its recommendation from case decision 2021-005-FB-UA and case decision 2020-005-FB-UA that Meta should explain to users precisely what rule in a content policy they have violated.</p>	<p>Enforcement</p>	<p>Somewhat Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>
<p>Ayahuasca brew #3</p> <p>Q42021</p>	<p>To respect diverse traditional and religious expressions and practices, the Board recommends that Meta modify the Instagram Community Guidelines and Facebook Regulated Goods Community Standard to allow positive discussion of traditional and religious uses of non-medical drugs where there is historic evidence of such use. The Board also recommends that Meta make public all allowances, including existing allowances.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



<p>Alleged crimes in Raya Kobo #1</p> <p>Q42021</p>	<p>Meta should rewrite Meta’s value of “Safety” to reflect that online speech may pose risk to the physical security of persons and the right to life, in addition to the risks of intimidation, exclusion and silencing.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Alleged crimes in Raya Kobo #2</p> <p>Q42021</p>	<p>Facebook’s Community Standards should reflect that in the contexts of war and violent conflict, unverified rumors pose higher risk to the rights of life and security of persons. This should be reflected at all levels of the moderation process.</p>	<p>Content Policy</p>	<p>Not Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Alleged crimes in Raya Kobo #3</p> <p>Q42021</p>	<p>Meta should commission an independent human rights due diligence assessment on how Facebook and Instagram have been used to spread hate speech and unverified rumors that heighten the risk of violence in Ethiopia. The assessment should review the success of measures Meta took to prevent the misuse of its products and services in Ethiopia. The assessment should also review the success of measures Meta took to allow for corroborated and public interest reporting on human rights atrocities in Ethiopia. The assessment should review Meta’s language capabilities in Ethiopia and if they are adequate to protect the rights of its users. The assessment should cover a period from June 1, 2020, to the present. The company should complete the assessment within six months from the moment it responds to these recommendations. The assessment should be published in full.</p>	<p>Transparency</p>	<p>Somewhat Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>



Asking for Adderall #1	Q12022 Meta should publish its internal definitions for “non-medical drugs” and “pharmaceutical drugs” in the Facebook Community Standard on Restricted Goods and Services. The published definitions should: (a) make clear that certain substances may fall under either “non-medical drugs” or “pharmaceutical drugs” and (b) explain the circumstances under which a substance would fall into each of these categories. The Board will consider this recommendation implemented when these changes are made in the Community Standard.	Content Policy	Comprehensive	Implementation demonstrated through published information
Asking for Adderall #2	Q12022 Meta should study the consequences and trade-offs of implementing a dynamic prioritization system that orders appeals for human review, and consider whether the fact that an enforcement decision resulted in an account restriction should be a criterion within this system. The Board will consider this recommendation implemented when Meta shares the results of these investigations with the Board and in its quarterly Board transparency report.	Enforcement	Comprehensive	Progress reported



<p>Asking for Adderall #3</p>	<p>Q12022</p> <p>Meta should conduct regular assessments on reviewer accuracy rates focused on the Restricted Goods and Services policy. The Board will consider this recommendation implemented when Meta shares the results of these assessments with the Board, including how these results will inform improvements to enforcement operations and policy development, and summarize the results in its quarterly Board transparency reports . Meta may consider if these assessments should be extended to reviewer accuracy rates under other Community Standards.</p>	<p>Enforcement</p>	<p>Not Comprehensive</p>	<p>Progress reported</p>
<p>Swedish journalist reporting sexual violence against minors #1</p>	<p>Q12022</p> <p>Meta should define graphic depiction and sexualization in the Child Sexual Exploitation, Nudity and Abuse Community Standard. Meta should make clear that not all explicit language constitutes graphic depiction or sexualization and explain the difference between legal, clinical or medical terms and graphic content. Meta should also provide a clarification for distinguishing child sexual exploitation and reporting on child sexual exploitation. The Board will consider the recommendation implemented when language defining key terms and the distinction has been added to the Community Standard.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



Swedish journalist reporting sexual violence against minors #2

Q12022

Meta should undergo a policy development process, including as a discussion in the Policy Forum, to determine whether and how to incorporate a prohibition on functional identification of child victims of sexual violence in its Community Standards. This process should include stakeholder and expert engagement on functional identification and the rights of the child. The Board will consider this recommendation implemented when Meta publishes the minutes of the Product Policy Forum where this is discussed.

Content Policy

Comprehensive

Progress reported

Sharing private residential information PAO #1

Q12022

removing the “publicly available” policy exception. Meta should remove the exception that allows the sharing of private residential information (both images that currently fulfill the Privacy Violations policy’s criteria for takedown and addresses) when considered “publicly available”. This means Meta would no longer allow otherwise violating content on Facebook and on Instagram if “published by at least five news outlets” or if it contains residential addresses or imagery from financial records or statements of an organization, court records, professional and business licenses, sex offender registries or press releases from government agencies, or law enforcement. The Board will consider this implemented when Meta modifies its Internal Implementation Standards and its content policies.

Content policy

Comprehensive

Progress reported



Sharing private residential information PAO #2

Q12022

Making sure the newsworthiness exception is consistently applied. Meta should develop and publicize clear criteria for content reviewers to escalate for additional review of public interest content that potentially violates the Community Standards but may be eligible for the newsworthiness exception, as previously recommended in case decision 2021-010-FB-UA. These criteria should cover, in addition to large protests as highlighted in Recommendation No. 3 from case decision 2021-010-FB- UA, content that shares information on investigations of corruption, tax evasion, money laundering and other criminal activities, in particular where property has been attained through such activities. The Board will consider this implemented when Meta publicly shares these escalation criteria.

Content Policy

Not Comprehensive

Recommendation omitted, declined, or reframed



Sharing private residential information PAO #3

Q12022

keeping public interest information in the form of images on the platform [this does not apply to addresses]. Meta should allow the sharing of “imagery that displays the external view of private residences” when the property depicted is the focus of the news story, even when the following conditions listed in the Privacy Violations Community Standard are met (“the residence is a single-family home, or the resident's unit number is identified in the image/caption”; “the city/neighborhood or GPS pin” is identified; “the content identifies the resident(s)”; and “that same resident objects to the exposure of their private residence”). However, Meta should not allow the sharing of such information when there is a “context of organizing protests against the resident,” that is, an attempt to organize protests in the future, and not news reporting on protests that have already taken place. The Board will consider this implemented when Meta modifies its content policies.

Content Policy

Comprehensive

Progress reported

Sharing private residential information PAO #4

Q12022

Allowing the organization of protests at publicly owned official residences. Meta should allow the publication of addresses and imagery of official residences provided to high-ranking government officials, such as heads of state, heads of federal or local government, ambassadors and consuls. The Board will consider this implemented when Meta modifies its content policies.

Content Policy

Somewhat Comprehensive

Progress reported



<p>Sharing private residential information PAO #5</p> <p>Q12022</p>	<p>Strengthening the role of user consent. Meta should allow the resharing of private residential addresses when posted by the affected user themselves or when the user consented to its publication. Users should not be presumed to consent to private information posted by others. The Board will consider this implemented when Meta modifies its content policies.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Sharing private residential information PAO #6</p> <p>Q12022</p>	<p>Strengthening the role of user consent. Users should have a quick and effective mechanism to request the removal of private information posted by others. We will consider this implemented when Meta demonstrates in its transparency reports that user requests to remove their information are consistently and promptly actioned. This recommendation is not applicable to official residences of high-ranking government officials.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation</p>
<p>Sharing private residential information PAO #7</p> <p>Q12022</p>	<p>Improving the clarity of the rules. Meta should better explain, in the text of Facebook’s Privacy Violations policy, when disclosing the city where a residence is located will suffice for the content to be removed, and when disclosing its neighborhood would be required for the same matter (e.g., by specifically referencing the population threshold at which sharing only the city as part of the content will no longer be considered violating). The Board will consider this implemented when Meta modifies its content policies.</p>	<p>Content Policy</p>	<p>Somewhat Comprehensive</p>	<p>Progress reported</p>



<p>Sharing private residential information PAO #8</p>	<p>Q12022</p> <p>Improving the clarity of the rules. Meta should explain, in the text of Facebook’s Privacy Violations policy, its criteria for assessing whether the resident is sufficiently identified in the content. The Board will consider this implemented when Meta modifies its content policies.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Sharing private residential information PAO #9</p>	<p>Q12022</p> <p>Improving the clarity of the rules. The Board reiterates Recommendation No. 1 from case decision 2021-013-IG-UA that Meta should explain to users that it enforces the Facebook Community Standards on Instagram, with several specific exceptions. The Board notes Meta’s response to these recommendations. While Meta may be taking other actions to comply with the recommendations, the Board recommends Meta update the introduction to the Instagram Community Guidelines (“The Short” Community Guidelines) within 90 days to inform users that if content is considered violating on Facebook, it is also considered violating on Instagram, as stated in the company’s Transparency Center, with some exceptions. Meta should also provide a link to the Privacy Violations Community Standard in the language of the Community Guidelines. The Board will consider this implemented when Meta modifies its content policies.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



Sharing private residential information PAO #10	Q12022	Responding more quickly to situations of increased risk. Meta should let users reporting content that may violate the Privacy Violations policy provide additional context about their claim. The Board will consider this implemented when Meta publishes information about its appeal processes that demonstrate users may provide this context in appeals.	Enforcement	Somewhat Comprehensive Progress reported
Sharing private residential information PAO #11	Q12022	Responding more quickly to situations of increased risk. Meta should create a specific channel of communications for victims of doxing (available both for users and non-users) that may be easily accessed, allow the victim to explain in detail their situation and risks the content creates for them, and prompt swift action from the company. Additionally, Meta could provide financial support to organizations that already have hotlines in place. Meta should prioritize action when the impacted person references belonging to a group facing heightened risk to safety in the region where the private residence is located. The Board will consider this implemented when Meta creates the channel and publicly announces how to use it.	Enforcement	Somewhat Comprehensive Recommendation omitted, declined, or reframed



Sharing private residential information PAO #12

Q12022

Responding more quickly to situations of increased risk. Meta should consider the violation of its Privacy Violations policy as “severe,” prompting temporary account suspension, in cases where the sharing of private residential information is clearly related to malicious action that created a risk of violence or harassment. The Board will consider this implemented when Meta updates its Transparency Center description of the strikes system to make clear that some Privacy Violations are severe and may result in account suspension.

Enforcement

Somewhat Comprehensive

Progress reported



Sharing private residential information PAO #13

Q12022

Reversing enforcement errors. Meta should give users an opportunity to remove or edit private information within their content following a removal for violation of the Privacy Violations policy (i.e., a threat of harm is present, but the story could be allowed on the platform if the image of residence or other information, such as the address, the city/neighborhood, GPS pin, or the name and picture of the resident was removed). The company could issue a notification of the violation and give the user a short deadline for them to act on the content, during which the content should be made temporarily unavailable. If the user removes/edits the private residential information out of the content within the deadline, the temporary block on the content would be lifted. If not, then the content would remain blocked. The Board will consider this implemented when Meta publishes information about its enforcement processes that demonstrates users are notified of specific policy violations when content is removed and granted a remedial window before the content is permanently deleted.

Enforcement

Comprehensive

Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation

Sharing private residential information PAO #14

Q12022

Reversing enforcement errors. Meta should let users indicate in their appeals against content removal that their content falls into one of the exceptions to the Privacy Violations policy. The Board will consider this implemented when Meta publishes information about its appeal processes that demonstrates users may provide this

Enforcement

Somewhat Comprehensive

Progress reported



<p>Sharing private residential information PAO #15</p> <p>Q12022</p>	<p>information in appeals.</p> <p>Improving enforcement transparency. Meta should publish quantitative data on the enforcement of the Privacy Violations policy in the company's Community Standards Enforcement Report. The Board will consider this implemented when Meta's transparency report includes Privacy Violations enforcement data.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Sharing private residential information PAO #16</p> <p>Q12022</p>	<p>Improving enforcement transparency. Meta should break down data in its transparency reports to indicate the amount of content removed following privacy-related government requests, even if taken down under the Privacy Violations policy and not under local privacy laws. The Board will consider this implemented when Meta's transparency reporting includes all government requests that result in content removal for violating the Privacy Violations policy as a separate category.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>



<p>Sharing private residential information PAO #17</p> <p>Q12022</p>	<p>Improving enforcement transparency. Meta should provide users with more detail on the specific policy of the Privacy Violations Community Standard that their content was found to violate and implement it across all working languages of the company's platforms. The Board will consider this implemented when Meta publishes information and data about user notifications.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>
<p>Sudan graphic video #1</p> <p>Q22022</p>	<p>Meta should amend the Violent and Graphic Content Community Standard to allow videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses. This content should be allowed with a warning screen so that people are aware that content may be disturbing. The Board will consider this recommendation implemented when Meta updates the Community Standard</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Sudan graphic video #2</p> <p>Q22022</p>	<p>Meta should undertake a policy development process that develops criteria to identify videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses. The Board will consider this recommendation implemented when Meta publishes the findings of the policy development process, including information on the process and criteria for identifying this content at scale.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



Sudan Graphic Video #3	Q22022	Meta should make explicit in its description of the newsworthiness allowance all the actions it may take (for example, restoration with a warning screen) based on this policy. The Board will consider this recommendation implemented when Meta updates the policy.	Content Policy	Comprehensive Implementation demonstrated through published information
Sudan graphic video #4	Q22022	To ensure users understand the rules, Meta should notify users when it takes action on their content based on the newsworthiness allowance including the restoration of content or application of a warning screen. The user notification may link to the Transparency Center explanation of the newsworthiness allowance. The Board will consider this implemented when Meta rolls out this updated notification to users in all markets and demonstrates that users are receiving this notification through enforcement data.	Enforcement	Comprehensive Progress reported



<p>Reclaiming Arabic words #1</p> <p>Q22022</p>	<p>Meta should translate the Internal Implementation Standards and Known Questions to Modern Standard Arabic. Doing so could reduce over-enforcement in Arabic-speaking regions by helping moderators better assess when exceptions for content containing slurs are warranted. The Board notes that Meta has taken no further action in response to the recommendation in the "Myanmar Bot" case (2021-007-FB-UA) that Meta should ensure that its Internal Implementation Standards are available in the language in which content moderators review content. The Board will consider this recommendation implemented when Meta informs the Board that translation to Modern Standard Arabic is complete.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>
<p>Reclaiming Arabic words #2</p> <p>Q22022</p>	<p>Meta should publish a clear explanation on how it creates its market-specific slur lists. This explanation should include the processes and criteria for designating which slurs and countries are assigned to each market-specific list. The Board will consider this implemented when the information is published in the Transparency Center.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Partial Implementation demonstrated through published information</p>



<p>Reclaiming Arabic words #3</p> <p>Q22022</p>	<p>Meta should publish a clear explanation of how it enforces its market-specific slur lists. This explanation should include the processes and criteria for determining precisely when and where the slurs prohibition will be enforced, whether in respect to posts originating geographically from the region in question, originating outside but relating to the region in question, and/or in relation to all users in the region in question, regardless of the geographic origin of the post. The Board will consider this recommendation implemented when the information is published in Meta’s Transparency Center.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Reclaiming Arabic words #4</p> <p>Q22022</p>	<p>Meta should publish a clear explanation on how it audits its market-specific slur lists. This explanation should include the processes and criteria for removing slurs from or keeping slurs on Meta's market-specific lists. The Board will consider this recommendation implemented when the information is published in Meta’s Transparency Center.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Knin cartoon #1</p> <p>Q22022</p>	<p>Meta should clarify the Hate Speech Community Standard and the guidance provided to reviewers, explaining that even implicit references to protected groups are prohibited by the policy when the reference would reasonably be understood. The Board will consider this recommendation implemented when Meta updates its Community Standards and Internal Implementation Standards to content reviewers to incorporate this revision.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



Knin cartoon #2	Q22022	In line with Meta's commitment following the "Wampum belt" case (2021-012-FB-UA), the Board recommends that Meta notify all users who have reported content when, on subsequent review, it changes its initial determination. Meta should also disclose the results of any experiments assessing the feasibility of introducing this change with the public. The Board will consider this recommendation implemented when Meta shares information regarding relevant experiments and, ultimately, the updated notification with the Board and confirms it is in use in all languages.	Enforcement	Somewhat Comprehensive	Progress reported
Colombian police cartoon #1	Q32022	To improve Meta's ability to remove non-violating content from banks programmed to identify or automatically remove violating content, Meta should ensure that content with high rates of appeal and high rates of successful appeal is re-assessed for possible removal from its Media Matching Service banks. The Board will consider this recommendation implemented when Meta: (i) discloses to the Board the rates of appeal and successful appeal that trigger a review of Media Matching Service-banked content, and (ii) confirms publicly that these reassessment mechanisms are active for all its banks that target violating content.	Enforcement	Somewhat Comprehensive	Progress reported



<p>Colombian police cartoon #2</p> <p>Q32022</p>	<p>To ensure that inaccurately banked content is quickly removed from Meta’s Media Matching Service banks, Meta should set and adhere to standards that limit the time between when banked content is identified for re-review and when, if deemed non-violating, it is removed from the bank. The Board will consider this recommendation implemented when Meta: (i) sets and discloses to the Board its goal time between when a re-review is triggered and when the non-violating content is restored, and (ii) provides the Board with data demonstrating its progress in meeting this goal over the next year.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Colombian police cartoon #3</p> <p>Q32022</p>	<p>To enable the establishment of metrics for improvement, Meta should publish the error rates for content mistakenly included in Media Matching Service banks of violating content, broken down by each content policy, in its transparency reporting. This reporting should include information on how content enters the banks and the company’s efforts to reduce errors in the process. The Board will consider this recommendation implemented when Meta includes this information in its Community Standards Enforcement Report.</p>	<p>Transparency</p>	<p>Comprehensive</p>	<p>Recommendation omitted, declined, or reframed</p>



<p>Mention of the Taliban in news reporting #1</p>	<p>Q32022</p> <p>Meta should investigate why the December 2021 changes to the Dangerous Individuals and Organizations policy were not updated within the target time of six weeks, and ensure such delays or omissions are not repeated. The Board asks Meta to inform the Board within 60 days of the findings of its investigation, and the measures it has put in place to prevent translation delays in future.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Implementation demonstrated through published information</p>
<p>Mention of the Taliban in news reporting #2</p>	<p>Q32022</p> <p>Meta should make its public explanation of its two-track strikes system more comprehensive and accessible, especially for “severe strikes.” It should include all policy violations that result in severe strikes, which account features can be limited as a result and specify applicable durations. Policies that result in severe strikes should also be clearly identified in the Community Standards, with a link to the “Restricting Accounts” explanation of the strikes system. The Board asks Meta to inform the Board within 60 days of the updated Transparency Center explanation of the strikes system, and the inclusion of the links to that explanation for all content policies that result in severe strikes.</p>	<p>Content Policy</p>	<p>Comprehensive</p>	<p>Progress reported</p>



<p>Mention of the Taliban in news reporting #3</p>	<p>Q32022</p> <p>Meta should narrow the definition of “praise” in the Known Questions guidance for reviewers, by removing the example of content that “seeks to make others think more positively about” a designated entity by attributing to them positive values or endorsing their actions. The Board asks Meta to provide the Board within 60 days with the full version of the updated Known Questions document for Dangerous Individuals and Organizations.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Mention of the Taliban in news reporting #4</p>	<p>Q32022</p> <p>Meta should revise its internal Implementation Standards to make clear that the “reporting” allowance in the Dangerous Individuals Organizations policy allows for positive statements about designated entities as part of the reporting, and how to distinguish this from prohibited “praise.” The Known Questions document should be expanded to make clear the importance of news reporting in situations of conflict or crisis and provide relevant examples, and that this may include positive statements about designated entities like the reporting on the Taliban in this case. The Board asks Meta to share the updated Implementation Standards with the Board within 60 days.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>



<p>Mention of the Taliban in news reporting #5</p>	<p>Q32022</p> <p>Meta should assess the accuracy of reviewers enforcing the reporting allowance under the Dangerous Individuals and Organizations policy in order to identify systemic issues causing enforcement errors. The Board asks Meta to inform the Board within 60 days of the detailed results of its review of this assessment, or accuracy assessments Meta already conducts for its Dangerous Individuals and Organizations policy, including how the results will inform improvements to enforcement operations, including for HIPO.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Mention of the Taliban in news reporting #6</p>	<p>Q32022</p> <p>Meta should conduct a review of the HIPO ranker to examine if it can more effectively prioritize potential errors in the enforcement of allowances to the Dangerous Individuals and Organizations Policy. This should include examining whether the HIPO ranker needs to be more sensitive to news reporting content, where the likelihood of false-positive removals that impacts freedom of expression appears to be high. The Board asks Meta to inform the Board within 60 days of the results of its review and the improvements it will make to avoid errors of this kind in the future.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>
<p>Mention of the Taliban in news reporting #7</p>	<p>Q32022</p> <p>Meta should enhance the capacity allocated to HIPO review across languages to ensure that more content decisions that may be enforcement errors receive additional human review. The Board asks Meta to inform the Board within 60 days of the planned capacity enhancements.</p>	<p>Enforcement</p>	<p>Comprehensive</p>	<p>Progress reported</p>

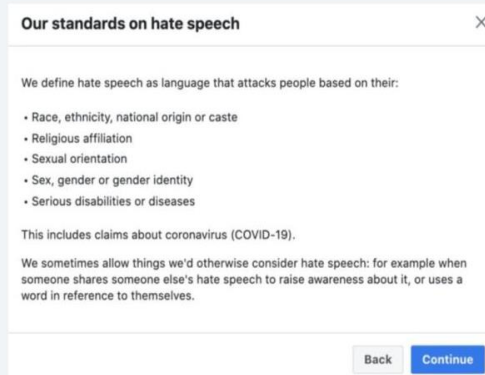


Annex III - Draft language for more granular notifications, shared by Meta with the Board

Product Mocks

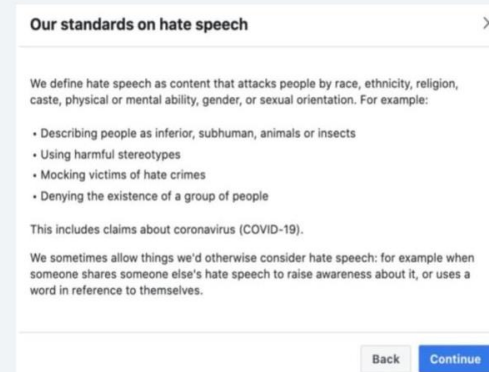
Hate Speech

Before: Anyone whose content was taken down for violating our Hate Speech policy received the same general message, which broadly referred to the standards of the policy.



Old – General **hate speech** policy

Now: Following on the board's rec, the messaging is now tailored to each subtype of hate speech violation (e.g. slurs; dehumanization). This improves transparency among our users and gives them a better idea of why their content was violating.



New – Specific to **dehumanization** rule

Languages: English, Spanish, Portuguese, Arabic, Hindi, Indonesian

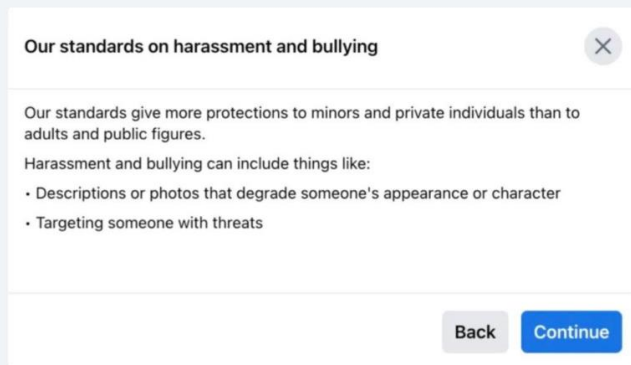


Product Mocks

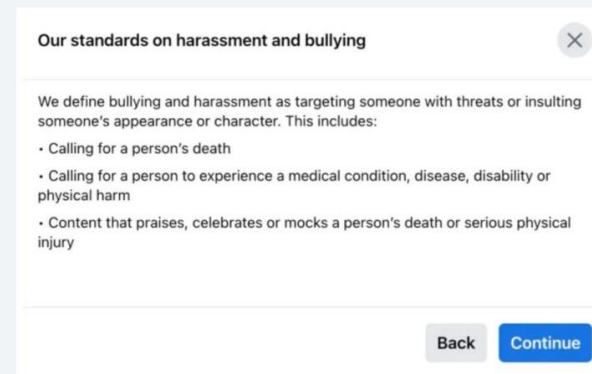
Bullying & Harassment

Before: Anyone whose content was taken down for violating our Bullying & Harassment policy received the same general message, which broadly refers to the standards of the policy.

Now: Although B&H policy doesn't have subtype violations as the Hate Speech policy does, following the board's rec we were able to improve and better specify what violates our policy, which also resulted to improvements in Legitimacy.



Old – General **bullying & harassment** policy



New – **Bullying & harassment** with more detail

Languages: English, French, Hindi, Spanish, Arabic, Portuguese and Indonesian