



# **Oversight Board transparency reports - Q4 2020, Q1 & Q2 2021.**



## Table of contents

1. Executive summary – p3
2. Glossary of terms – p10
3. Special report on Facebook’s cross-check system – p11
4. Transparency report for fourth quarter of 2020 – p16
5. Transparency report for first quarter of 2021 – p24
6. Transparency report for second quarter of 2021 – p47
7. Annex – Summaries of cases where Facebook identified its original decision on a piece of content was incorrect – p71



## 1. Executive summary

The transparency reports in this document cover the fourth quarter of 2020, as well as the first and second quarters of 2021. These include data on cases submitted to the Board, as well as further detail on our decisions and recommendations. Moving forward, the Board will publish a report with data about its work as soon as possible after the end of each quarter.

Here are six important points from our first transparency reports, covering October 22, 2020 – June 30, 2021:

### 1. Over half a million user appeals submitted

Between October 2020 and the end of June 2021, Facebook and Instagram users had submitted around 524,000 cases to the Board. User appeals increased in each quarter, with around 114,000 cases in Q4 2020, 203,000 cases in Q1 2021, and 207,000 in Q2 2021. Facebook also submitted 35 cases.

In total, the Board selected 21 cases to review and ultimately proceeded with 17 of these. By the end of June, the Board had decided 11 cases – overturning Facebook’s decision eight times and upholding it three times.

On average it took 74 days to decide and implement these cases. The Board’s Bylaws require this to happen within 90 days, apart from in exceptional circumstances.

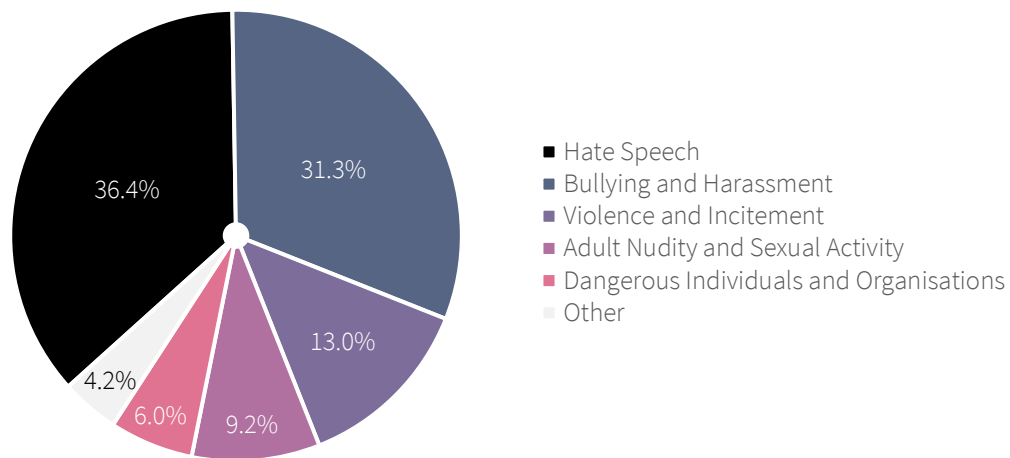
As part of these decisions, the Board made 52 recommendations to Facebook and received 9,842 public comments. The vast majority of these comments (9,666) related to the case on former US President Donald Trump.



## 2. Two thirds of appeals where users wanted their content restored related to hate speech or bullying

Up to the end of June, we estimate that more than a third of cases submitted to the Board (36%) related to content concerning Facebook’s rules on Hate Speech.

We estimate that Bullying and Harassment made up another third (31%) of cases submitted, with Violence and Incitement (13%), Adult Nudity and Sexual Activity (9%) and Dangerous Individuals and Organizations (6%) making up most of the remaining cases. These figures do not include user appeals to remove content from Facebook, which were introduced beginning in mid-April.

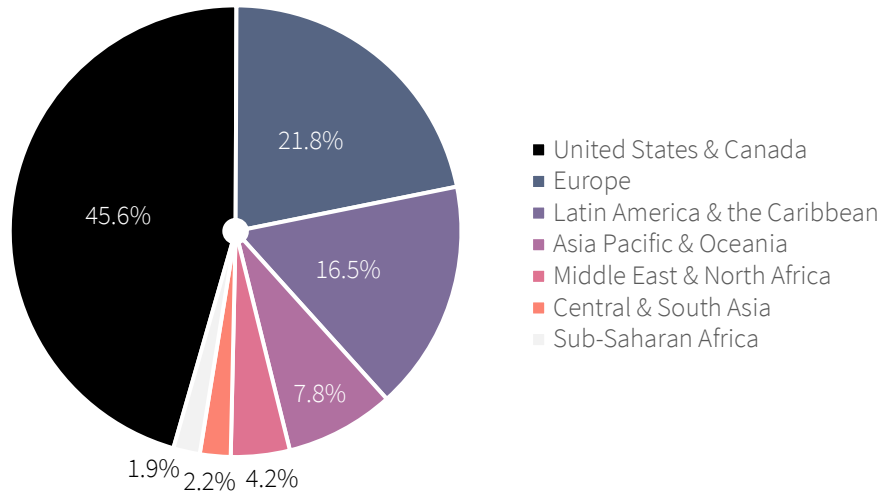


## 3. Nearly half of user appeals came from the United States & Canada

Up to the end of June, we estimate that nearly half of cases submitted (46%) came from the US & Canada, while 22% of cases came from Europe and 16% from Latin America and the Caribbean. We estimate that 8% of cases came from the Asia Pacific & Oceania region, 4% came from the Middle East and North



Africa, 2% came from Central and South Asia and 2% came from Sub-Saharan Africa.



We do not believe this represents the actual distribution of Facebook content issues around the globe. If anything, we have reason to believe that users in Asia, Sub-Saharan Africa, and the Middle East experience more, not fewer, problems with Facebook than parts of the world with more appeals.

We are expanding our outreach in these areas to ensure that Board oversight extends to users everywhere, and we ask that users and civil society organizations in Asia, Sub-Saharan Africa and the Middle East take notice of our concern and bring appeals when they suffer the effects of poor content moderation by Facebook in their areas.

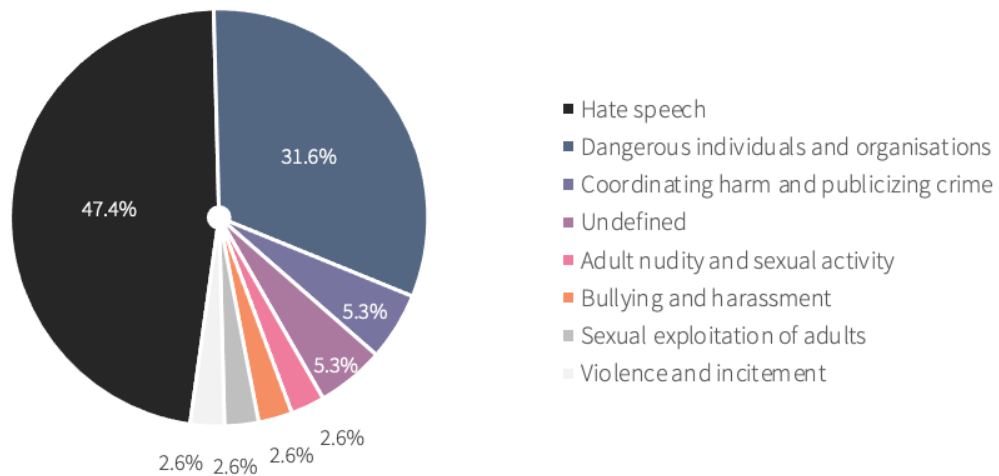
#### **4. The Board’s wider processes prompted Facebook to restore more than 30 pieces of content covering significant cases**

As part of the process for shortlisting cases for consideration, the Board asks Facebook to confirm that cases are eligible for review under the Bylaws. As a



result of this process, by the end of June, Facebook identified 38 shortlisted cases where its original decision on a piece of content was incorrect.

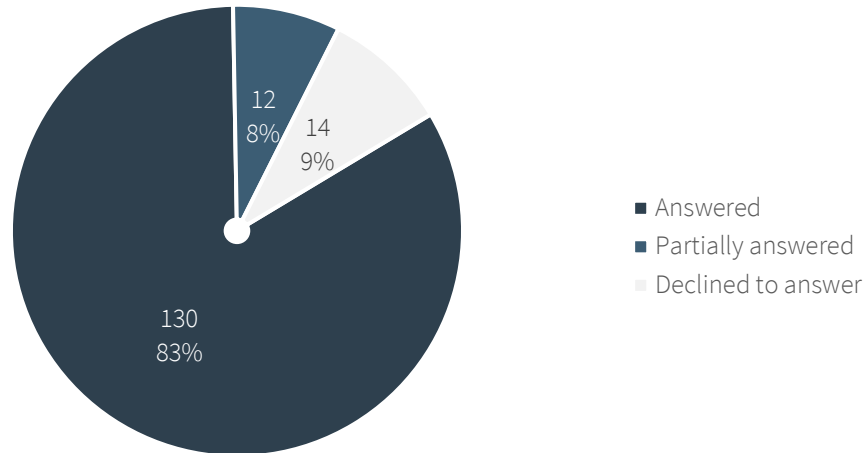
In 35 of these cases, Facebook then took action on the content, while in three cases it could not do so as the content had been deleted by the user. Nearly half of the cases where Facebook identified its original decision as incorrect related to the Hate Speech Community Standard, while nearly a third related to Dangerous Individuals and Organizations.



## 5. Facebook is answering most of the Board’s questions, but not all of them

To assist with making its decisions, the Board sends questions to Facebook about specific cases.

Of the 156 questions sent to Facebook about decisions we published by the end of June, Facebook answered 130, partially answered 12 and declined to answer 14.



## 6. Facebook was not fully forthcoming with the Board on cross-check

Following recent disclosures in the Wall Street Journal, [the Board committed](#) to look at whether Facebook had been forthcoming in its responses on its cross-check system, which the company uses to review content decisions relating to high-profile users.

In the Board’s view, the team within Facebook tasked with providing information has not been fully forthcoming on cross-check. On some occasions, Facebook failed to provide relevant information to the Board, while in other instances, the information it did provide was incomplete.

When Facebook referred [the case related to former US President Trump](#) to the Board, it did not mention the cross-check system. Given that the referral included a specific policy question about account-level enforcement for political leaders, many of whom the Board believes were covered by cross-check, this omission is not acceptable. Facebook only mentioned cross-check to the Board when we asked whether Mr. Trump’s page or account had been subject to ordinary content moderation processes.

In its subsequent briefing to the Board, Facebook admitted it should not have



said that cross-check only applied to a “small number of decisions.” Facebook noted that for teams operating at the scale of millions of content decisions a day, the numbers involved with cross-check seem relatively small, but recognized its phrasing could come across as misleading.

We also noted that Facebook’s response to our recommendation to “clearly explain the rationale, standards and processes of [cross-check] review, including the criteria to determine which pages and accounts are selected for inclusion” provided no meaningful transparency on the criteria for accounts or pages being selected for inclusion in cross-check.

The credibility of the Oversight Board, our working relationship with Facebook, and our ability to render sound judgments on cases all depend on being able to trust that information provided to us by Facebook is accurate, comprehensive, and paints a full picture of the topic at hand. We will continue to track and report on information provided by Facebook to ensure it is as comprehensive and complete as possible.

Today, the Board has announced it has accepted a request from Facebook, in the form of a policy advisory opinion, to review the company’s cross-check system and make recommendations on how it can be changed.

Specifically, Facebook requested guidance on, among other things: how to ensure fairness and objectivity in cross-check reviews, taking into account context; how to govern cross-check and promote transparency; and the criteria it uses to determine who is included in cross-check and how to ensure this is equitable.

Now that we have accepted Facebook’s request, the Board will engage with civil society globally, including academics and researchers, as we scrutinize this critical issue. This will include a call for public comments which we will launch in the coming days. The Board continues to reach out to a broad range of voices to inform its work, including former Facebook employees who have come forward in recent months.





Facebook has now agreed to share with the Board documents concerning cross-check as reported on in the Wall Street Journal. The Board will review these as we produce our policy advisory opinion.

Facebook has also agreed that, from now on, it commits to provide information about the wider context which may be relevant to the Board's case decisions. This should give a fuller understanding of the work Facebook has already done on a given topic. We will include analysis on whether Facebook is fulfilling this commitment in our future transparency reporting.

Once the Board has deliberated on this policy advisory opinion, and voted to approve it, we will issue our recommendations to Facebook. Facebook must then respond within 30 days.

## **What's next**

From now on, we will be publishing transparency reports after each quarter ends. We will also issue an annual report which assesses Facebook's performance in implementing our decisions and recommendations in due course. We will publish all of these on our website.



## 2. Glossary of terms

**Annual report** – A report published 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 Facebook 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, comprised of at least five Oversight Board Members with membership rotating every three months, 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 or Community Guidelines).

**Facebook-referred case** – A case submitted to the Oversight Board by Facebook. Facebook has the ability to refer cases to the board both on an ongoing basis and under emergency circumstances, with the latter being heard under the process for expedited review.

**Facebook’s legal review** – Step in case selection process where Facebook 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).

**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.



### 3. Special report on Facebook’s cross-check system

Last month, new information emerged in the Wall Street Journal on Facebook’s cross-check system. These disclosures have drawn renewed attention to the seemingly inconsistent way that the company makes decisions and why greater transparency and oversight of Facebook matter so much for users.

This section looks at whether Facebook has been fully forthcoming to the Board on cross-check, including the practice of whitelisting. It includes information provided in a briefing by Facebook on cross-check requested by the Board following the disclosures, as well as an analysis of our decisions related to cross-check and Facebook’s responses to our recommendations.

In the Board’s view, the team within Facebook tasked to provide information has not been fully forthcoming in its responses on cross-check. On some occasions, Facebook failed to provide relevant information to the Board, while in other instances, the information it did provide was incomplete.

The credibility of the Oversight Board, our working relationship with Facebook, and our ability to render sound judgments on cases all depend on being able to trust that information provided to us by Facebook is accurate, comprehensive, and paints a full picture of the topic at hand. We will continue to track and report on information provided by Facebook to ensure it is as comprehensive and complete as possible.

While the briefing Facebook provided on cross-check provided some initial answers, the Board has additional questions and requires further information. Facebook has now agreed to share with the Board documents concerning cross-check as reported on in the Wall Street Journal. The Board will review these documents as we produce our policy advisory opinion on cross-check, and we have now accepted Facebook’s request for this opinion. Facebook will continue to brief the Board on cross-check and answer our questions, which will also feed into our policy advisory opinion.

Facebook has also agreed that, from now on, it commits to provide information about the wider context which may be relevant to the Board’s case decisions. This should give a fuller understanding of the work Facebook has already done on a given topic. We will include analysis on whether Facebook is fulfilling this commitment in our future transparency reporting.

- Which of the Board’s decisions first dealt with cross-check?

In May 2021, the Oversight Board published [its decision on the suspension of former US President Donald Trump’s accounts](#), upholding Facebook’s decision but finding that it failed to impose a proper penalty.

While the Board’s decision focused on Facebook’s cross-check system, the company did not mention cross-check when it referred this case to the Board, or in the initial case file materials it provided. Given that the referral included a specific policy question about account-level enforcement for political leaders, many of whom the Board now believes were covered by cross-check, it is not acceptable that Facebook did not mention cross-check in the information it initially provided. This was clearly relevant to the



Board’s deliberations and its ability to make informed recommendations. Facebook only mentioned the cross-check program in a response to one of the Board’s questions in the Trump case, which asked whether Mr. Trump’s page or account had been subject to ordinary content moderation processes. Facebook’s answer was consistent with a [2018 Facebook newsroom article](#) on cross-check (itself a response to a previous media leak) which Facebook cited and hyperlinked in its response. It was on the basis of this information that the Board addressed cross-check in its decision.

- **What has the Board said about cross-check in its decisions?**

The Trump decision raised concerns about a lack of transparency around cross-check and the different outcomes which might arise from applying different processes to different users.

In the decision, the Board wrote:

*Facebook told the Board it applies a “cross check” system to some “high profile” accounts to “minimize the risk of errors in enforcement.” For these accounts, Facebook sends content found to violate its Community Standards for additional internal review. After this escalation, Facebook decides if the content is violating. Facebook told the Board that “it has never had a general rule that is more permissive for content posted by political leaders.” While the same general rules apply, the “cross check” system means that decision-making processes are different for some “high profile” users.*

While the Board did not explicitly support or oppose the idea of cross-check as it was presented by Facebook, the Trump decision included guidance at odds with the cross-check program as described by the Wall Street Journal, particularly the idea that “the same rules should apply to all users of the platform” (Recommendations, Section 10). The decision also expressed concerns that “different processes may lead to different substantive outcomes” (Rationale, Section 8.1).

The Board also noted in its Trump decision that there is “limited public information on cross-check” and that “the lack of transparency regarding these decision-making processes appears to contribute to perceptions that the company may be unduly influenced by political or commercial considerations” (Rationale, Section 8.1).

- **What did the Board say in its recommendations about cross-check and how did Facebook respond?**

In response to these concerns, the Board made two specific recommendations on cross-check in its Trump decision.

1. **First recommendation on cross-check**

The Board recommended that “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.”

In [Facebook’s response](#), the company claimed it was “implementing fully” the Board’s recommendation (which it numbered ‘No. 12’). Facebook said “Our Community Standards apply around the world to all



types of content and are designed so they can be applied consistently and fairly to a community that transcends regions, cultures, and languages. Today we are providing more information about our system of reviews for public figures' content, which includes our cross check process and newsworthiness allowance, in our Transparency Center.”

While this recommendation was not the focus of the Wall Street Journal's reporting, the [Transparency Center explanation](#) of cross-check is not detailed. The explanation provides a limited rationale for cross-check focused on “accuracy” and “consistency,” which the Wall Street Journal's reporting brings into question as the program's main purpose.

The new Transparency Center explanation provided also states that cross-check is for “high-visibility content,” implying that the program applies to individual pieces of content based on reach, rather than the identity of the account or page. The fact that Facebook provided such an ambiguous, undetailed response to a call for greater transparency is not acceptable. Facebook's answer provides no meaningful transparency on the criteria for accounts or pages being selected for inclusion in cross-check, which was at the heart of the Board's recommendation.

In the Board's briefing with Facebook, the company noted that, while there is no single list of users who get cross-check review, teams across Facebook put together a list of entities with a high risk of “false positive” content moderation decisions (i.e., erroneous “take down” decisions). Facebook noted that they completed an average of under 10,000 cross-check reviews per day. The company also said that, currently, users aren't informed when they are enrolled in cross-check. The Board has also learned that users whose appeals the Board selects are added to cross-check.

## **2. Second recommendation on cross-check**

The Board also recommended that “**Facebook should report on the relative error rates and thematic consistency of determinations made through the cross-check process compared with ordinary enforcement procedures.**”

In [Facebook's response](#), the company said it would take “no further action” on this recommendation (which it numbered ‘No. 13’). Facebook said that “while the Board has requested details about the relative error rates of enforcement decisions made through cross-check, we do not have systems in place to make this comparison. Our measurement accuracy systems are not designed to review the *small number of decisions* made through the cross check process” (emphasis added).

Facebook's claim that its cross-check system only applies to a “small number of decisions” was central to the Wall Street Journal's reporting that the company misled the Board. The Wall Street Journal states that “Despite attempts to rein it in, XCheck grew to include at least 5.8 million users in 2020” (para. 7) and that “tens of thousands of accounts” were added “just last year” (para. 50). While the Board never directly asked Facebook to disclose the size of the cross-check program, the Trump decision states that “Facebook told the Board it applies a ‘cross check’ system to *some* ‘high profile’ accounts to ‘minimize the risk of errors in enforcement’” (Section 2, emphasis added).



In the Board’s briefing with Facebook, the company admitted that it should not have said cross-check only applied to a “small number of decisions.” Facebook noted that for teams operating at scales of millions of content decisions a day, the numbers involved with cross-check seem relatively small, but recognized its phrasing could come across as misleading.

The Wall Street Journal also claimed that an internal document reviewing the program in 2019 showed that, in practice, cross-check operated to exempt highly influential accounts from enforcement, with up to 90% of reports not being reviewed at all. In the Board’s briefing with Facebook, the company claimed that currently 84% of the content produced by entities in the cross-check system which is initially reported as violating is reviewed by cross-check moderators. In response, Board Members expressed concern that the cross-check system has such a sizeable backlog.

Facebook also claimed that the numbers published in the Wall Street Journal were based on root-cause analysis on anecdotal cases which Facebook routinely conducts to identify potential gaps, but do not reach Facebook’s usual rigor for reporting metrics externally. The Board requested to see this research, as well as any additional information providing context for why it is incorrect. The Board expects to be kept updated as the company invests in research that meets its standards of rigor. Board Members also noted that Facebook could have provided more information in their response to this recommendation. Even if Facebook could not provide definitive error rates, it could have noted that some initial research had been conducted, and shared some of the results in the interests of transparency.

The Wall Street Journal indicates that when cross-check review takes place, Facebook is interpreting its rules in favor of content remaining on the platform. It claims that the incentives for review teams to avoid “PR fires” appears to be considerable, and that this leads to the same rules being interpreted differently to allow content from influential accounts to remain on the platform when it would likely be removed for a regular user.

Among the cases the Board has decided so far, we have already encountered contradictory situations linked to cross-check. In the ‘Shared Al Jazeera post’ decision, a verified Al Jazeera page, which was subject to cross-check, was able to benefit from a policy exception for reporting on dangerous individuals and organizations. However, a user who was not subject to cross-check and shared Al Jazeera’s post found their content removed from the platform, despite a similar exception in the policy for “neutral discussion,” which permits the sharing of news reporting. This example highlights the Board’s concern, expressed in the Trump decision, that “different processes may lead to different substantive outcomes.” It also highlights the real impacts on users, and the importance of Facebook providing more information on the relative error rates between ordinary procedures and those for cross-check. It is currently unclear to what extent Facebook has a higher tolerance for errors affecting ordinary users compared to those enrolled in cross-check. A number of the Board’s decisions have highlighted the frequency with which regular users have their content wrongly removed, often resulting in account suspensions or deletions.

- **The practice of “whitelisting”**

The Wall Street Journal article also claimed the existence of a practice of “whitelisting” where certain accounts or pages have been exempted from content enforcement. The Board notes with concern that



the practical consequence of having low levels of review through cross-check results in effective “whitelisting” of accounts or pages, or at least delayed enforcement against potentially dangerous content. In telling the Board that cross-check involves sending content for additional review, Facebook failed to explain that some percentage of this content is not reviewed and remains on the platform despite its initial assessment as violating Facebook policies. This is a good example of how less-than-forthcoming answers from Facebook deprived the Board of critical information and context. However, the article also refers to a “political whitelist” that meant leeway was given to prominent political accounts. It alleges that some “VIP users” have also been provided with a “self-remediation window” to delete content themselves. The Board notes that, in the Trump decision, Facebook refused to answer one of the Board’s questions about whether the company had been contacted by political officeholders or their staff about the suspension of Mr. Trump’s accounts.

In the Board’s briefing with Facebook, the company defined “whitelisting” as the exemption of certain accounts or pages from enforcement of the Community Standards. Facebook denied that it engages in this practice, noting that there is only one set of Community Standards which applies to all users, including those enrolled in cross-check. The company denied that the same policies are applied more leniently to cross-checked accounts and pages. This response is in line with Facebook’s response to one of the Board’s questions in the Trump case which claimed that “Cross check does not exempt any account or content from Facebook’s Community Standards or Instagram’s Community Guidelines. Rather, it is used to ensure that non-violating content is not inadvertently removed.” As we learn more about cross-check, the Board will continue to probe whether it functions to exclude certain entities from content moderation.

- **Policy advisory opinion on cross-check**

On September 21, following disclosures in the Wall Street Journal, the Board publicly called on Facebook to [commit to maximum transparency on its cross-check system](#). The next day, Facebook held a briefing with the Board on cross-check, where Board Members reiterated this call and urged Facebook to involve the Board more closely in key policy questions like cross-check.

On September 28, Facebook [announced it would be requesting](#) a policy advisory opinion from the Board on cross-check. Specifically, Facebook requested guidance on, among other things: (i.) how to ensure fairness and objectivity in cross-check reviews, taking into account context; (ii.) how to govern cross-check and promote transparency; and the criteria it uses to determine who is included in cross-check and how to ensure this is equitable.

Now that we have accepted Facebook’s request, the Board will engage with civil society globally, including academics and researchers as we scrutinize this critical issue. This will include a call for public comments which we will launch in the coming days. The Board continues to reach out to a broad range of voices to inform its work, including former Facebook employees who have come forward in recent months. Once the Board has deliberated on this policy advisory opinion, and voted to approve it, we will issue our recommendations to Facebook publicly. Facebook must then respond within 30 days.



## **4. Transparency Report for fourth quarter of 2020**

This transparency report for the fourth quarter of 2020 (October 1 – December 31, 2020) sets out key statistics on cases selected by the Board.

In the fourth quarter of 2020, the Board began its work in earnest. On October 22, 2020, the Oversight Board began accepting cases from users and Facebook, while on December 1, 2020 the Board announced the selection of its first cases.





## Key Findings – Fourth quarter 2020<sup>1</sup>



~114,428 cases submitted to the Oversight Board, of which seven were submitted by Facebook.

27 user-submitted cases longlisted.

27 user-submitted cases shortlisted.

Seven cases assigned to panel, of which two were submitted by Facebook.

No cases decided.

<sup>1</sup> 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 Q1 but longlisted in Q2 would appear in “submitted cases” in Q1 and “longlisted cases” in Q2. The next footnote provides further context on the numbers of submitted cases.



## Q4 Submitted User Cases

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

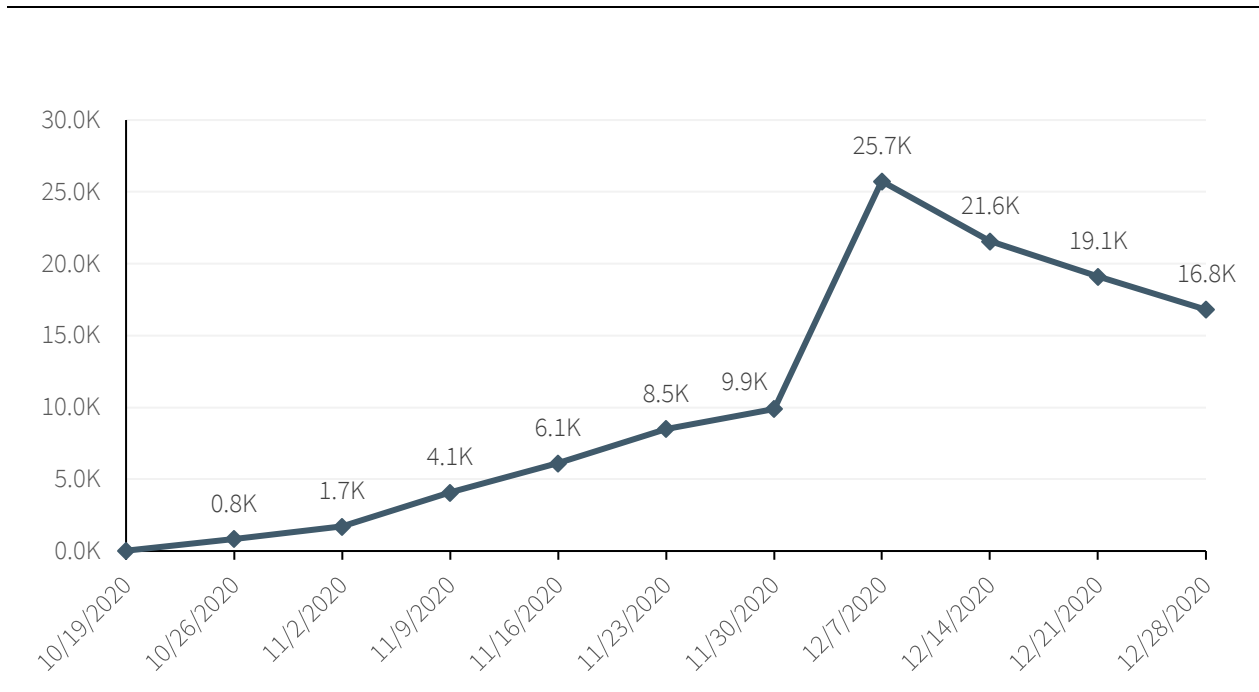
In this quarter, due to limitations in the functionality of the Case Management Tool (CMT), submitted cases have been counted manually by the Case Selection Team as they were seen in CMT at the time. As such, these numbers should be taken as an estimate and preliminary.

In this period, an estimated **114,421 cases** were submitted by users.

The number of cases submitted to the Oversight Board has increased significantly since the Board started accepting cases on October 22, 2020. In the week of December 7, 2020, submitted cases jumped from 9,900 per week to 25,700 per week after the Oversight Board announced its initial round of cases and received significant media coverage.

### Estimated number of cases submitted to Oversight Board by week

Number of cases, thousands

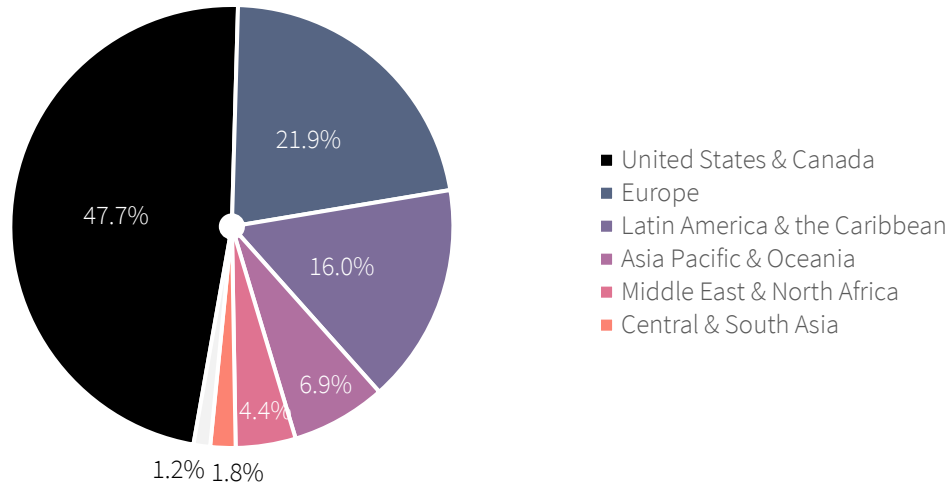




The majority of submitted cases were from the United States & Canada (47.7%) and Europe (21.9%).

### Estimated cases submitted by user-selected region

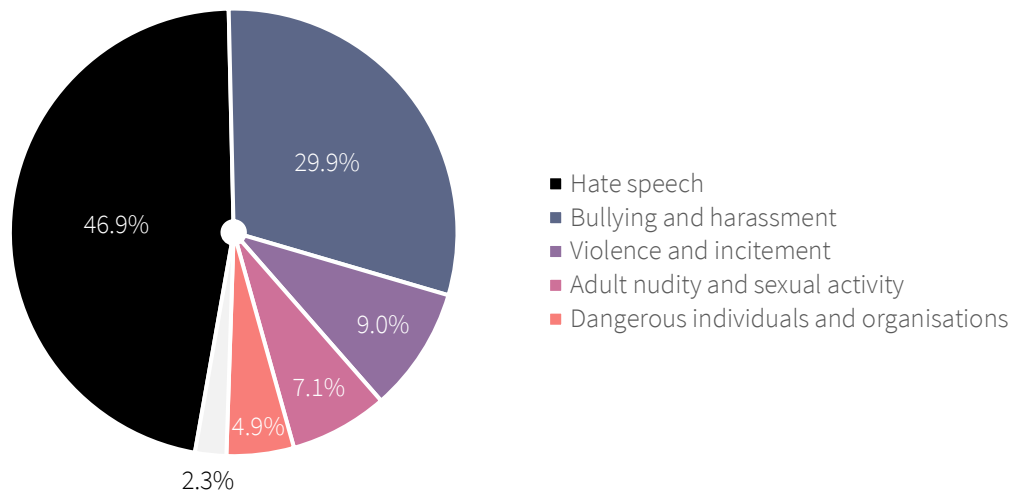
Percent<sup>2</sup>



Users primarily submitted cases concerning Facebook’s Hate Speech (46.9%) and Bullying and Harassment (29.9%) policies. Other notable policies of dispute were Violence and Incitement (9.0%), Adult Nudity and Sexual Activity (7.1%) and Dangerous Individuals and Organizations (4.9%).

### Estimated cases submitted by Community Standard

Percent<sup>3</sup>



<sup>2</sup> The distribution is approximate and based on a sample of 81,272 cases. Throughout this report, numbers may not add up to 100.0% due to rounding.

<sup>3</sup> The distribution is approximate and based on a sample of 44,633 cases.



## Q4 Longlisted User Cases

In this period, **27 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 26 pieces of content on Facebook and one piece of content on Instagram. The cases involved 18 posts and nine comments.

33.3% of these cases were from Europe and 25.9% from the United States & Canada. Cases were longlisted from 17 countries. The United States had the highest number of longlisted cases.

### Longlisted cases by region and country

Number of cases

Europe	9
Armenia	1
Azerbaijan	2
France	2
Italy	2
Portugal	1
Ukraine	1
United States & Canada	7
United States	7
Asia Pacific & Oceania	5
Japan	1
Malaysia	1
Singapore	1
Taiwan	2
Central & South Asia	2
Afghanistan	1
India	1
Sub-Saharan Africa	2
Gambia	1
Nigeria	1
Middle East & North Africa	1
Israel	1
Latin America & Caribbean	1
Brazil	1
<b>Total</b>	<b>27</b>

Almost half of the cases longlisted (44.4%) concerned Facebook’s Hate Speech policy. Cases concerning Dangerous Individuals and Organizations (22.2%) were also highly represented.



### Longlisted cases by Community Standard

Number of cases

Hate Speech	12
Dangerous Individuals and Organizations	6
Violence and Incitement	4
Adult Nudity and Sexual Activity	3
Bullying and Harassment	1
Sexual Exploitation of Adults	1
<b>Total</b>	<b>27</b>

## Q4 Shortlisted User Cases

The Case Selection Committee identifies a shortlist of cases from the longlist to consider for selection. Board Members on the Case Selection Committee rotate every three months, evaluating and selecting cases. In this quarter, the Case Selection Committee met four times and **shortlisted 27 user-submitted cases**.

The shortlist is sent to Facebook’s legal team to review for legal obligations, as per the Bylaws. Such obligations may mean a case is 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, Facebook has also assessed whether its original decision on a piece of content was correct or not, including deciding to restore content wrongly removed. This does not affect the eligibility of the content under review.

Of the 27 cases shortlisted in this period, all were confirmed as eligible by Facebook’s legal team.

A response on one case was only received from Facebook’s legal team several weeks after the deadline agreed by Facebook and the Board. Although no legal obligations were identified, and the case was eventually approved, the case did not proceed to the final shortlist due to this reason.

In this quarter, Facebook determined that nine of the cases shortlisted by the Oversight Board were wrongful applications of its Community Standards. In eight of these cases, Facebook then reversed its original decision and restored the content, while in one case the content had been deleted by the user.

### Cases where Facebook 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	5	0	5
Adult Nudity and Sexual Activity	0	1	1
Dangerous Individuals and Organizations	1	0	1
Sexual Exploitation of Adults	1	0	1
Violence and Incitement	1	0	1
<b>Total</b>	<b>8</b>	<b>1</b>	<b>9</b>



## Q4 Submitted Facebook Cases

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

Facebook submitted seven cases to the Oversight Board during this period. Referred cases fell under Facebook’s Violence and Incitement, Bullying and Harassment, and Hate Speech Community Standards.

### Facebook referrals considered<sup>4</sup>

<i>Case ID</i>	<i>Name</i>	<i>Facebook’s decision</i>	<i>Platform</i>	<i>Language</i>	<i>Community Standard</i>	<i>Countries<sup>5</sup></i>
2020-006-FB-FBR	Claimed COVID cure	Take down	FB	French	Violence and Incitement	France
2020-007-FB-FBR	Protest in India against France	Take down	FB	Hindi/English	Violence and Incitement	India & France
N/A	N/A	Keep up	FB	English	N/A	United States
N/A	N/A	Take down	FB	French	Violence and Incitement	France
N/A	N/A	Take down	FB	English	Bullying and Harassment	Fiji
N/A	N/A	Take down	FB	French	Hate Speech	France
N/A	N/A	Keep up	FB	English	N/A	United States

<sup>4</sup> Cases which are not selected for assignment do not have a Case ID.

<sup>5</sup> 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.



## Q4 Assigned Cases

The Case Selection Committee assigns cases to panels. In this period, the Case Selection Committee assigned seven cases to panel.

### Cases assigned

<i>Case ID</i>	<i>Name</i>	<i>Date</i>	<i>Platform</i>	<i>Source</i>	<i>Community Standard</i>	<i>Countries<sup>6</sup></i>
2020-001-FB-UA	Former Malaysian Prime Minister tweets	11/30/20	FB	User	Hate Speech	Malaysia & France
2020-002-FB-UA	Myanmar post about Muslims	11/30/20	FB	User	Hate Speech	Myanmar, France & China
2020-003-FB-UA	Armenians in Azerbaijan	11/30/20	FB	User	Hate Speech	Armenia & Azerbaijan
2020-004-IG-UA	Breast cancer symptoms and nudity	11/30/20	IG	User	Adult Nudity and Sexual Activity	Brazil
2020-005-FB-UA	Nazi quote	11/30/20	FB	User	Dangerous Individuals and Organizations	United States
2020-006-FB-FBR	Claimed COVID cure	11/30/20	FB	FB	Violence and Incitement	France
2020-007-FB-FBR	Protest in India against France	12/03/20	FB	FB	Violence and Incitement	India & France

## Q4 Decided Cases

In this period, no cases were decided. Shortly after case 2020-001-FB-UA was assigned, the post it was associated with was deleted by a user and the Board did not proceed with the case.

<sup>6</sup> 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.



## **5. Transparency Report for first quarter of 2021**

This transparency report for the first quarter of 2021 (January 1 – March 31, 2021) sets out key statistics on cases selected by the Board, as well as the decisions and recommendations we made in this quarter.

In the first quarter of 2021, the Board issued its first decisions and recommendations, and accepted a case regarding former US President Donald Trump.





## Key Findings – First quarter 2021<sup>7</sup>



~202,740 cases submitted to the Oversight Board, of which 19 were submitted by Facebook.

92 user-submitted cases longlisted.

39 user-submitted cases shortlisted.

Five cases assigned to panel, of which one was submitted by Facebook.

Six cases decided, with 18 recommendations for Facebook.

<sup>7</sup> 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 Q1 but longlisted in Q2 would appear in “submitted cases” in Q1 and “longlisted cases” in Q2. The next footnote provides further context on the numbers of submitted cases.



## Q1 Submitted User Cases

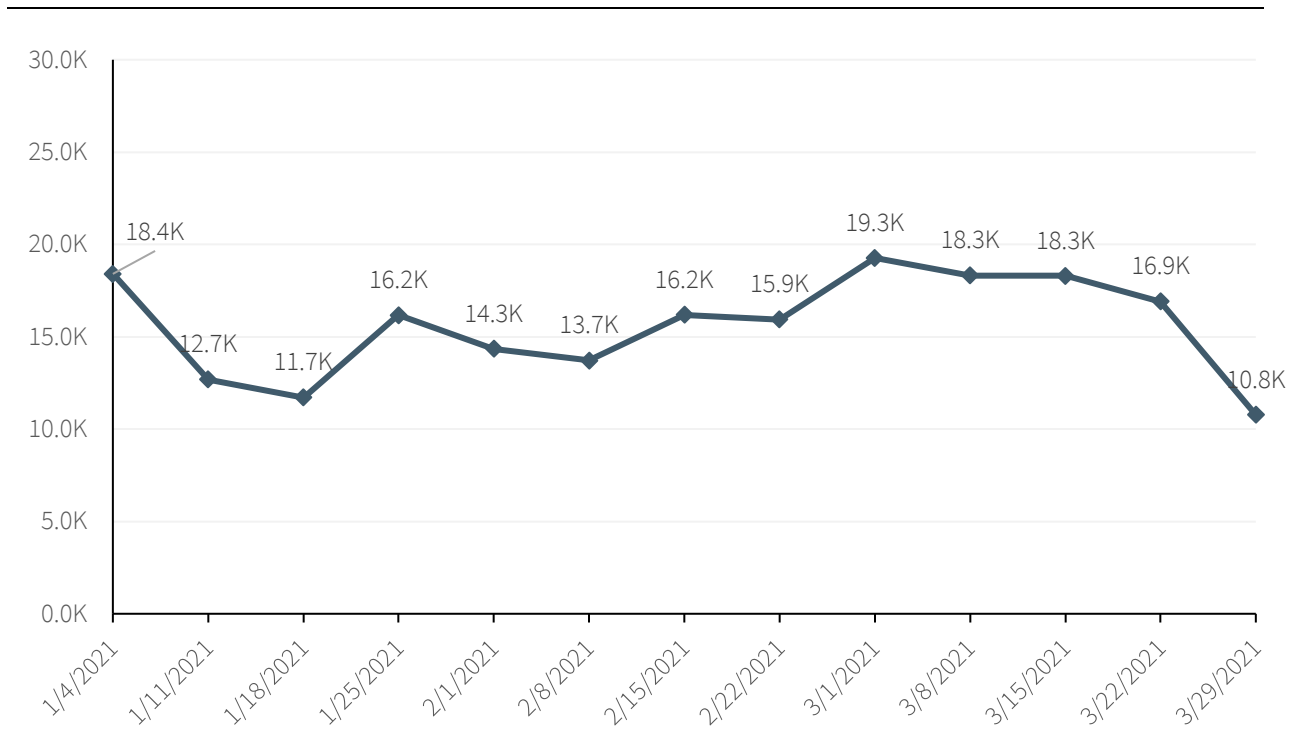
Where users have exhausted Facebook’s appeals process, they can challenge the company’s decision by appealing eligible content to the Oversight Board.

In this quarter, due to limitations in the functionality of the Case Management Tool (CMT), submitted cases have been counted manually by the Case Selection Team as they were seen in CMT at the time. As such, these numbers should be taken as an estimate and preliminary.

In this period, an estimated **202,721 cases** were submitted by users.

### Estimated number of cases submitted to Oversight Board by week

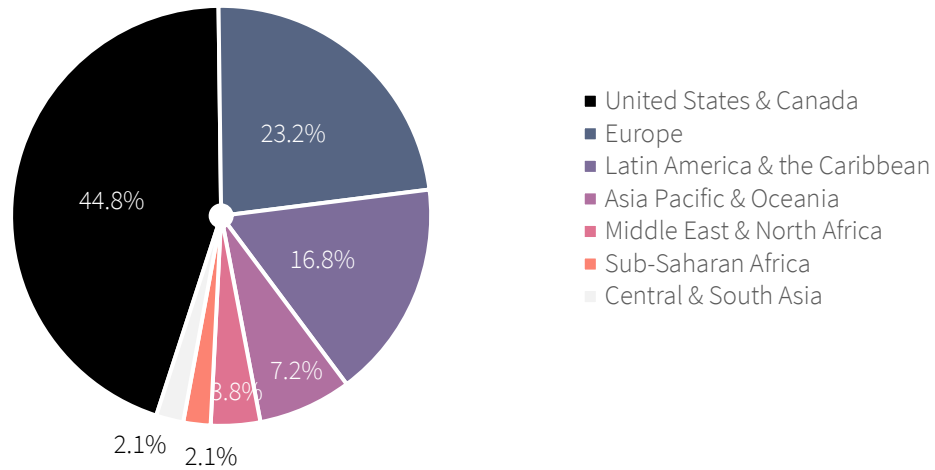
Number of cases, thousands





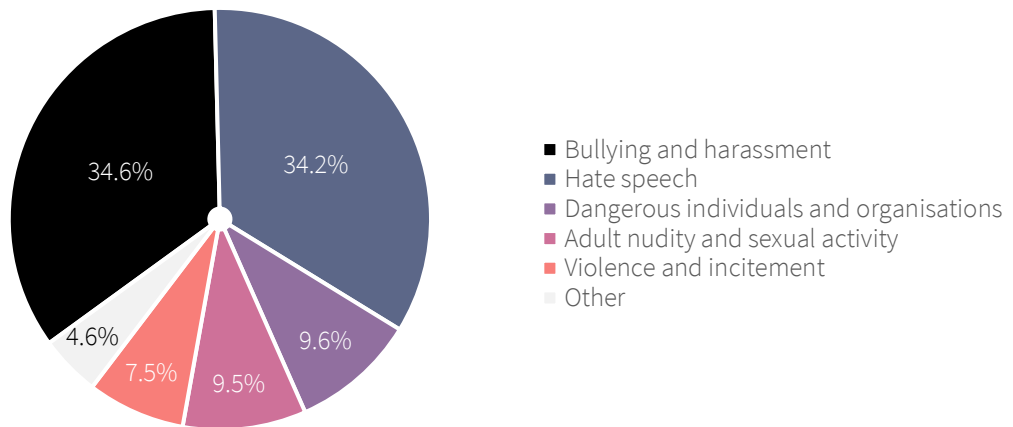
The majority of submitted cases were from the United States & Canada (44.8%) and Europe (23.2%).

### Estimated cases submitted by user-selected region Percent<sup>8</sup>



Users primarily submitted cases concerning Facebook’s Bullying and Harassment (34.6%) and Hate Speech (34.2%) policies. Other notable policies of dispute were Dangerous Individuals and Organizations (9.6%), Adult Nudity and Sexual Activity (9.5%), and Violence and Incitement (7.5%).

### Estimated cases submitted by Community Standard Percent<sup>9</sup>



<sup>8</sup> The distribution is approximate and based on a sample of 155,864 cases. Throughout this report, numbers may not add up to 100.0% due to rounding.

<sup>9</sup> The distribution is approximate and based on a sample of 155,355 cases.



## Q1 Longlisted User Cases

In this period, **92 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 72 pieces of content on Facebook and 20 pieces of content on Instagram. The cases involved 75 posts and 17 comments.

26.1% of the cases were from the United States & Canada, and 18.5% from Europe. Cases were longlisted from 42 countries. The United States, India, and Brazil had the highest number of longlisted cases.

### Longlisted cases by region and country<sup>10</sup>

Number of cases

<b>United States &amp; Canada<sup>11</sup></b>	<b>24</b>
Canada	1
United States <sup>11</sup>	23
<b>Europe</b>	<b>17</b>
Belgium	1
Bulgaria	1
Cyprus	1
France	1
Germany	2
Macedonia	1
Netherlands	1
Romania	1
Russia	3
Spain	1
Turkey	1
Ukraine	1
United Kingdom	2
<b>Central &amp; South Asia</b>	<b>14</b>
Afghanistan	2
Bangladesh	1
India	6
Kazakhstan	1
Myanmar	1
Pakistan	1
Sri Lanka	2
<b>Middle East &amp; North Africa</b>	<b>11</b>

<sup>10</sup> 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.

<sup>11</sup> One case relates to both Malaysia and the United States.



Iran	4
Iraq	3
Israel	1
Lebanon	1
Syria	1
Yemen	1
<b>Latin America &amp; Caribbean</b>	<b>10</b>
Argentina	1
Brazil	6
Chile	1
Colombia	1
Trinidad and Tobago	1
<b>Asia Pacific &amp; Oceania<sup>11</sup></b>	<b>9</b>
China	1
Indonesia	1
Malaysia <sup>11</sup>	2
Singapore	3
South Korea	1
Taiwan	1
<b>Sub-Saharan Africa</b>	<b>8</b>
Ethiopia	4
Nigeria	2
South Africa	2
<b>Total</b>	<b>92</b>

About a third of the cases longlisted (36.9%) concerned Facebook’s Hate Speech policy. Cases concerning Dangerous Individuals and Organizations (30.4%), as well as Adult Nudity and Sexual Activity (8.7%), were also highly represented.

#### Longlisted cases by Community Standard

Number of cases

Hate Speech	34
Dangerous Individuals and Organizations	28
Adult Nudity and Sexual Activity	8
Bullying and Harassment	6
Coordinating Harm and Publicizing Crime	5
Violence and Incitement	4
Cruel and Insensitive	3
Sexual Exploitation of Adults	2
Suicide and Self-Injury	1
Regulated Goods	1
<b>Total</b>	<b>92</b>



## Q1 Shortlisted User Cases

The Case Selection Committee identifies a shortlist of cases from the longlist to consider for selection. Board Members on the Case Selection Committee rotate every three months, evaluating and selecting cases by majority vote.

In this quarter, the Case Selection Committee reviewed and shortlisted cases on five occasions. While its members generally meet every two weeks, this cadence has been modified when there is no panel with the capacity to accept new cases. In this period, the committee **shortlisted 39 cases**.

The shortlist is sent to Facebook’s legal team to review for legal obligations, as per the Bylaws. Such obligations may mean a case is 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, Facebook has also assessed whether its original decision on a piece of content was correct or not, including deciding to restore content wrongly removed. This does not affect the eligibility of the content under review.

Of the 39 cases shortlisted in this period, 34 were confirmed as eligible by Facebook’s legal team and five cases were removed from the Board’s review either by the person who appealed the content or through Facebook’s legal review.

In four of these five cases, the user deleted the content during the appeals process. One case was excluded by Facebook’s legal team as ineligible under the Bylaws.

The case was excluded from the Board’s review for reasons related to the Dangerous Individuals and Organizations Community Standard.

In this quarter, Facebook determined that 17 of the cases shortlisted by the Oversight Board were wrongful applications of Facebook’s Community Standards (See Annex). In 15 of these cases, Facebook then reversed its original decision and restored the content, while in two cases the content had been deleted by the user.

### Cases where Facebook 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	7	1	8
Dangerous Individuals and Organizations	4	3	7
Coordinating Harm and Publicizing Crime	2	0	2
<b>Total</b>	<b>13</b>	<b>4</b>	<b>17</b>



## Q1 Submitted Facebook Cases

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

Facebook submitted 19 cases to the Oversight Board during this period. 13 were considered during this period and six were carried over for consideration in the next period. Referred cases fell under Facebook’s Dangerous Individuals and Organizations, Violence and Incitement and Bullying and Harassment Community Standards.

### Facebook referrals considered<sup>12</sup>

<i>Case ID</i>	<i>Name</i>	<i>Facebook’s decision</i>	<i>Platform</i>	<i>Language</i>	<i>Community Standard</i>	<i>Countries<sup>13</sup></i>
2021-001-FB-FBR	Former President Trump’s suspension	Take down	FB	English	Dangerous Individuals and Organizations	United States
N/A	N/A	Take down	FB	English	Violence and Incitement	United States
N/A	N/A	Take down	FB	English	Violence and Incitement	United States
N/A	N/A	Keep up	FB	English	N/A	United States
N/A	N/A	Keep up	FB	English	N/A	United States
N/A	N/A	Keep up	IG	English	N/A	Ireland

<sup>12</sup> Cases which are not selected for assignment do not have a Case ID.

<sup>13</sup> 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.



N/A	N/A	Keep up	IG	Russian	N/A	Kazakhstan
N/A	N/A	Keep up	FB	Chinese	N/A	China
N/A	N/A	Keep up	FB	English	N/A	United Kingdom
N/A	N/A	Take down	FB	Arabic	Dangerous Individuals and Organizations	Israel
N/A	N/A	Keep up	FB	Latvian	N/A	Latvia
N/A	N/A	Keep up	IG	German	N/A	Turkey
N/A	N/A	Take down	FB	Polish	Bullying and Harassment	Poland

## Q1 Assigned Cases

The Case Selection Committee assigns cases to panels. In this period, the Case Selection Committee assigned five cases to panel.

### Cases assigned

<i>Case ID</i>	<i>Name</i>	<i>Date</i>	<i>Platform</i>	<i>Source</i>	<i>Community standard</i>	<i>Countries<sup>14</sup></i>
2021-001-FB-FBR	Former President Trump's suspension	1/26/21	FB	FB	Dangerous Individuals and Organizations	United States

<sup>14</sup> 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.





2021-002-FB-UA	Depiction of Zwarte Piet	1/26/21	FB	User	Hate Speech	Netherlands
2021-003-FB-UA	Punjabi concern over the RSS in India	2/5/21	FB	User	Dangerous Individuals and Organizations	India
2021-004-FB-UA	Pro-Navalny protests in Russia	2/25/21	FB	User	Bullying and Harassment	Russia
2021-005-FB-UA	“Two buttons” meme	2/25/21	FB	User	Cruel and Insensitive / Hate Speech	Armenia, Turkey & United States



## Q1 Decided Cases

After being selected, the Board assigns cases to a five-member panel. Members of the panel are randomly chosen but include at least one member from the region implicated in the content and a mix of gender representation.

The panel looks at whether Facebook’s decision is consistent with its content policies and values, and its international human rights responsibilities. The Board’s decisions are binding, and Facebook must implement them within seven days of publication.

In this period, the Oversight Board **decided six cases**. The Oversight Board overturned Facebook’s decision five times and upheld Facebook’s decision once. **One case was not decided** as the post associated with it was deleted by a user.

### Cases decided

Case ID	Name	Platform	Source	Language of content	Community standard	Countries <sup>15</sup>	Outcome
2020-001-FB-UA	Former Malaysian Prime Minister tweets	FB	User	English	Hate Speech	Malaysia & France	Moot <sup>16</sup>
2020-002-FB-UA	Myanmar post about Muslims	FB	User	Burmese	Hate Speech	Myanmar, France & China	Overturned
2020-003-FB-UA	Armenians in Azerbaijan	FB	User	Russian	Hate Speech	Armenia & Azerbaijan	Upheld
2020-004-IG-UA	Breast cancer symptoms and nudity	IG	User	Portuguese	Adult Nudity and Sexual Activity	Brazil	Overturned
2020-005-FB-UA	Nazi quote	FB	User	English	Dangerous Individuals and Organizations	United States	Overturned

<sup>15</sup> 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.

<sup>16</sup> After this case was assigned, the post it was associated with was deleted by a user. Accordingly, no decision was made.



2020-006-FB-FBR	Claimed COVID cure	FB	FB	French	Violence and Incitement	France	Overturned
2020-007-FB-FBR	Protest in India against France	FB	FB	Hindi/English	Violence and Incitement	India & France	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.

**Human rights standards referenced**

Source	2020-002-FB-UA	2020-003-FB-UA	2020-004-IG-UA	2020-005-FB-UA	2020-006-FB-FBR	2020-007-FB-FBR
<b>UN Treaties</b>						
<b>ICCPR<sup>17</sup></b>						
Non-Discrimination & Remedies (Art. 2)	✓	✓	✓	✓		
Life (Art. 6)	✓	✓		✓		✓
Liberty and security of person (Art. 9)	✓	✓		✓		✓
Privacy (Art. 17)			✓			
Expression (Art. 19)	✓	✓	✓	✓	✓	✓
Incitement (Art. 20)	✓	✓		✓		✓
Equality (Art. 26)	✓	✓		✓		
<b>ICERD<sup>18</sup></b>						
Non-Discrimination (Art. 1)		✓		✓		
Hate Speech (Art. 4)		✓		✓		
Equality (Art. 5)		✓		✓		
<b>ICESCR<sup>19</sup></b>						
Health (Art. 12)			✓			
<b>CEDAW<sup>20</sup></b>						
Non-Discrimination (Art. 1)			✓			
<b>Convention on the Rights of the Child (CRC)</b>						
Life, Survival & Development (Art. 6)			✓			

<sup>17</sup> International Covenant on Civil and Political Rights  
<sup>18</sup> International Convention on the Elimination of All Forms of Racial Discrimination  
<sup>19</sup> International Covenant on Economic, Social and Cultural Rights  
<sup>20</sup> Convention on the Elimination of All Forms of Discrimination Against Women



## UN Treaty Bodies: Guidance & Recommendations

Human Rights Committee						
General Comment 31 on General Legal Obligation (CCPR/C/21/Rev.1/Add. 13)				✓		
General Comment 34 on the Freedoms of Opinion and Expression (CCPR/C/GC/34)	✓	✓	✓	✓	✓	✓
General Comment 35 on the Right to Liberty & Security of Person (CCPR/C/GC/35)					✓	
General Comment 36 on the Right to Life (CCPR/C/GC/36)		✓			✓	
Committee on Economic, Social & Cultural Rights						
General Comment 14 on the Right to Health (E/C.12/2000/4)				✓		
Committee on the Rights of the Child						
General Comment 13 on the Right to Freedom from Violence (CRC/C/GC/13)				✓		
Committee on the Elimination of Racial Discrimination						
General Recommendation 35 on Combatting Racist Hate Speech (CERD/C/GC/35)		✓			✓	
<b>Other UN Human Rights Standards</b>						
Responsibilities of Businesses						
Guiding Principles on Business and Human Rights (UNGPs)	✓	✓	✓	✓	✓	✓
UN Working Group on human rights & business enterprises, report on Conflict-Affected Regions (A/75/212)		✓				
Incitement						
Rabat Plan of Action (A/HRC/22/17/Add.4)	✓	✓		✓		✓
UN SR Freedom of Expression Reports						
Children (A/69/335)		✓		✓		
Artificial intelligence (A/73/348)		✓	✓	✓		
Online hate speech (A/74/486)	✓	✓				✓
Online content regulation (A/HRC/38/35)		✓	✓	✓	✓	
Disease pandemics (A/HRC/44/49)		✓	✓	✓	✓	
Elections in the digital age (research paper 1/2019)						✓
UN SR Racism Reports						
Neo-Nazism (A/HRC/38/53)						



## Decision timelines

Except for exceptional circumstances, the Bylaws specify that case decisions and implementation will occur a maximum of 90 days from the date the case is selected for review by the Case Selection Committee.<sup>21</sup> For the six decisions the Oversight Board published in this quarter, the average number of days from assignment of case to implementation of the Board’s decision has been 62 days.

<i>Case ID</i>	<i>Name</i>	<i>Beginning of 90-day period</i>	<i>Board’s decision published</i>	<i>Facebook implements decision</i>	<i>Number of days taken out of 90 days</i>
2020-001-FB-UA	Former Malaysian Prime Minister tweets	N/A	N/A	N/A	N/A
2020-002-FB-UA	Myanmar post about Muslims	11/30/20	1/28/21	1/28/21	59
2020-003-FB-UA	Armenians in Azerbaijan	11/30/20	1/28/21	N/A (Upheld FB’s decision)	59
2020-004-IG-UA	Breast cancer symptoms and nudity	11/30/20	1/28/21	N/A (FB already restored content)	59
2020-005-FB-UA	Nazi quote	11/30/20	1/28/21	1/28/21	59

<sup>21</sup> It became apparent as the Board started to accept cases that it would not be feasible in many instances to use the initially envisaged timeline for decisions (i.e. 90 days from the date of Facebook’s last decision on the case under review). As a result, the Bylaws were amended to provide for the typical 90-day period to start on the date when the Board assigned a case to panel. This update helped to ensure that all cases would have the same amount of time available for deliberation, no matter when the case was referred. The Board continues to refine its processes in light of the volume of appeals received, including ongoing review of timelines.



2020-006-FB-FBR	Claimed COVID cure	11/30/20	1/28/21	1/28/21	59
2020-007-FB-FBR	Protest in India against France	12/03/20	2/12/21	2/18/21	77

### Questions for Facebook

To assist with making its decisions, the Oversight Board sends questions to Facebook. Of the 43 questions sent by the Oversight Board to Facebook about decisions published in this quarter, Facebook answered 38 questions and declined to answer five questions.

Of the five questions which Facebook declined to answer, four concerned the user’s previous behavior on Facebook, which the company claimed was irrelevant to the Board’s determination about the case in hand.

One question requested that Facebook provide translations of the comments on a post. The company declined, claiming that it was not technically feasible to provide the information.

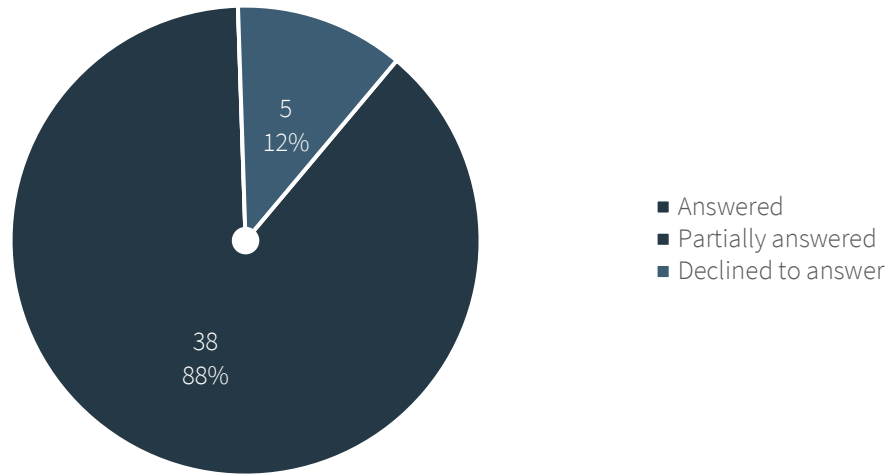
### Oversight Board questions answered by Facebook

Number of questions

<i>Case ID</i>	<i>Name</i>	<i>Answered</i>	<i>Partially answered</i>	<i>Declined to answer</i>	<i>Total</i>
2020-002-FB-UA	Myanmar post about Muslims	2	0	2	4
2020-003-FB-UA	Armenians in Azerbaijan	8	0	1	9
2020-004-IG-UA	Breast cancer symptoms and nudity	3	0	0	3
2020-005-FB-UA	Nazi quote	8	0	1	9
2020-006-FB-FBR	Claimed COVID cure	9	0	0	9
2020-007-FB-FBR	Protest in India against France	8	0	1	9
<b>Total</b>		<b>38</b>	<b>0</b>	<b>5</b>	<b>43</b>



## Oversight Board questions answered by Facebook



### Public Comments

The Oversight Board conducts a public comment process to assist it in its decision making. In this period, the Oversight Board received 96 comments, 84 of which were published.

#### Public comments received by publication status

Number of comments

Case ID	Name	Comments published	Comments not published (no consent)	Comments not published (violated terms)	Total	Comments unattributed <sup>22</sup>
2020-002-FB-UA	Myanmar post about Muslims	9	2	0	11	4
2020-003-FB-UA	Armenians in Azerbaijan	28	3	1	32	12
2020-004-IG-UA	Breast cancer symptoms and nudity	23	1	0	24	8
2020-005-FB-UA	Nazi quote	12	1	2	15	2
2020-006-FB-FBR	Claimed COVID cure	7	1	0	8	2
2020-007-FB-FBR	Protest in India against France	5	0	1	6	1
<b>Total</b>		<b>84</b>	<b>8</b>	<b>4</b>	<b>96</b>	<b>29</b>

<sup>22</sup> Unattributed comments are published comments with the author's name redacted by request.



The majority of public comments (72.9%) came from individuals, while a minority came from organizations (27.1%).

### 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>
2020-002-FB-UA	Myanmar post about Muslims	7	4	11
2020-003-FB-UA	Armenians in Azerbaijan	25	7	32
2020-004-IG-UA	Breast cancer symptoms and nudity	17	7	24
2020-005-FB-UA	Nazi quote	12	3	15
2020-006-FB-FBR	Claimed COVID cure	5	3	8
2020-007-FB-FBR	Protest in India against France	4	2	6
<b>Total</b>		<b>70</b>	<b>26</b>	<b>96</b>

Most public comments submitted came from the United States & Canada (61.4%) and Europe (27.1%).

### Public comments received by region

Number of comments

<i>Region</i>	<i>2020-002-FB-UA</i>	<i>2020-003-FB-UA</i>	<i>2020-004-IG-UA</i>	<i>2020-005-FB-UA</i>	<i>2020-006-FB-FBR</i>	<i>2020-007-FB-FBR</i>	<i>Total</i>
United States & Canada	5	24	11	11	4	4	59
Europe	5	6	8	4	3	0	26
Latin America & Caribbean	0	0	5	0	0	1	6
Asia Pacific & Oceania	1	0	0	0	1	1	3
Central & South Asia	0	2	0	0	0	0	2
Middle East & North Africa	0	0	0	0	0	0	0
Sub-Saharan Africa	0	0	0	0	0	0	0
<b>Total</b>	<b>11</b>	<b>32</b>	<b>24</b>	<b>15</b>	<b>8</b>	<b>6</b>	<b>96</b>





## Recommendations

In addition to providing decisions on appealed content, the Oversight Board also provided **18 policy recommendations** to Facebook, which the company responded to publicly within 30 days.

These included recommendations on content policy (clarification or changes to rules), enforcement (clarification or changes to how rules are applied), and transparency (on disclosure of information to the public).

Below, the Board lists all policy recommendations listed in case decisions during Q1. Facebook has not always directly responded to the Board’s recommendations in the same format that the Board set out in case decisions. Therefore, the Board has made its best effort to map Facebook’s commitments to the Board’s policy guidance as originally stated in each case, which sometimes means there are a different total number of recommendations per decision.

The Board has reproduced Facebook’s initial commitments made within the 30-day response window. Facebook continues to update their on-going progress to implement the recommendations. The Board will continue to monitor Facebook’s implementation efforts and include information about this monitoring in future transparency reporting.

### Oversight Board recommendations to Facebook

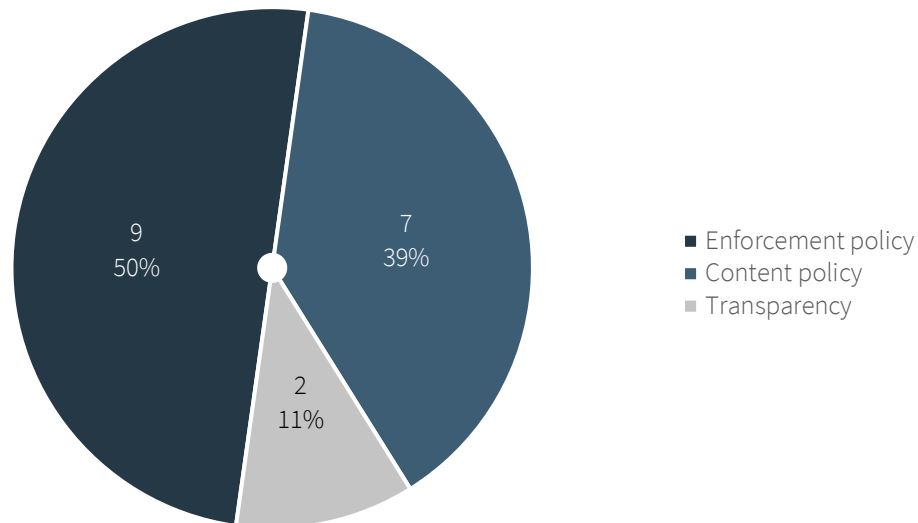
Number of recommendations

<i>Case ID</i>	<i>Name</i>	<i>Content policy</i>	<i>Enforcement</i>	<i>Transparency</i>	<i>Total</i>
2020-002-FB-UA	Myanmar post about Muslims	0	0	0	0
2020-003-FB-UA	Armenians in Azerbaijan	0	1	0	1
2020-004-IG-UA	Breast cancer symptoms and nudity	3	5	1	9
2020-005-FB-UA	Nazi quote	2	1	0	3
2020-006-FB-FBR	Claimed COVID cure	1	2	1	4
2020-007-FB-FBR	Protest in India against France	1	0	0	1
<b>Total</b>		<b>7</b>	<b>9</b>	<b>2</b>	<b>18</b>



## Oversight Board recommendations to Facebook

---



### Content policy recommendations

Recommendations to clarify rules, for rules to be modified, or for new rules to be created.

#### [Breast cancer symptoms and nudity \(2020-004-IG-UA\)](#)

- Revise the “short” explanation of the Instagram Community Guidelines to clarify that the ban on adult nudity is not absolute. **[No response]**<sup>23</sup>
- Revise the “long” explanation of the Instagram Community Guidelines to clarify that visible female nipples can be shown to raise breast cancer awareness. **[Facebook response (no. 2): “committed to action”]**
- Clarify that the Instagram Community Guidelines are interpreted in line with the Facebook Community Standards, and where there are inconsistencies the latter take precedence. **[Facebook response (no. 2): “committed to action”]**<sup>24</sup>

#### [Nazi quote \(2020-005-FB-UA\)](#)

- 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.

---

<sup>23</sup> Facebook merged three of the Board’s recommendations into one (Recommendation No. 2 in Facebook’s Transparency Center). However, the decision addressed both “the short” and “the long” aspects of Instagram’s Community Guidelines, and Facebook’s response only addressed the latter when describing the changes it would make.

<sup>24</sup> Facebook responded to a recommendation that was not contained in the policy advisory statement within the decision. The company “committed to action” to explain the relationship between Instagram Community Guidelines and Facebook’s Community Standards in user messaging when enforcing the policy. The Board has not counted this as a recommendation the Board made for the purposes of this transparency report.



The Community Standard should also better advise users on how to make their intent clear when discussing dangerous individuals or organizations. [Facebook response (no. 2): “committed to action”]

- 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. [Facebook response (no. 3): “assessing feasibility”]

#### [Claimed COVID cure \(2020-006-FB-FBR\)](#)

- 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). [Facebook responses (no. 1 and no. 6): “committed to action”]<sup>25</sup>

#### [Protest in India against France \(2020-007-FB-FBR\)](#)

- 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. [Facebook response (no. 1): “committed to action”]

#### **Enforcement recommendations**

Recommendations to change content moderation enforcement processes, including about the use of automated or human review, notification of enforcement action to users, and access to appeals.

#### [Armenians in Azerbaijan \(2020-003-FB-UA\)](#)

- 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. [Facebook response (no. 1): “assessing feasibility”]

#### [Breast cancer symptoms and nudity \(2020-004-IG-UA\)](#)

- 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. [Facebook response (no. 1): “committed to action”]

---

<sup>25</sup> Facebook split this recommendation in two for its response, and reworded parts which impacted the meaning substantially. The “commitment to action” therefore should be read as relating to Facebook’s reframed interpretation of the recommendation, rather than the Board’s original recommendation.



- 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. [No response]<sup>26</sup>
- Inform users when automation is used to take enforcement action against their content, including accessible descriptions of what this means. [Facebook response (no. 5): “assessing feasibility”]
- 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. [Facebook response (no. 4): “assessing feasibility”]<sup>27</sup>
- Implement an internal audit procedure to continuously analyze a statistically representative sample of automated content removal decisions to reverse and learn from enforcement mistakes. [No response]

#### [Nazi quote \(2020-005-FB-UA\)](#)

- 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). [Facebook response (no. 1): “committed to action”]<sup>28</sup>

#### [Claimed COVID cure \(2020-006-FB-FBR\)](#)

- 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;

---

<sup>26</sup> Facebook issued its responses to five decisions on one day. Facebook created a recommendation in its responses to this case that the Board did not make (see footnote 24). However, in explaining its commitments under that recommendation, it referenced its response to recommendations in “Armenians in Azerbaijan” 2020-003-FB-UA and “Nazi quote” 2020-005-FB-UA that are similar.

<sup>27</sup> The Board notes that Facebook only quoted the first line of this recommendation in their response. The remainder of this recommendation was not addressed.

<sup>28</sup> The Board notes that Facebook’s response to this recommendation did not address the final clause of this sentence.



- 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. [Facebook responses (no. 2, no. 3, and no. 4): “committed to action”]<sup>29</sup>
- 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. [Facebook response (no. 7): “no further action”]<sup>30</sup>

## Transparency

Recommendations to increase the transparency of Facebook’s content moderation, including through its transparency reports.

### [Breast cancer symptoms and nudity \(2020-004-IG-UA\)](#)

- 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. [Facebook response (no. 6): “assessing feasibility”]

### [Claimed COVID cure \(2020-006-FB-FBR\)](#)

- 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;

---

<sup>29</sup> Facebook’s response split this recommendation into three, and for each the response was “committed to action.”

<sup>30</sup> The Board’s decision in this case rested upon the Community Standard on Violence and Incitement, which uses the term “imminence” in relation to harm resulting from misinformation and unverifiable rumours. This recommendation addressed content that may cause harm *that is not imminent*, and therefore does not violate the Community Standard as drafted. Facebook’s response in this case indicates the Board’s recommendation for a policy change was not fully understood; the response instead focused on whether the imminence standard was met in this particular case, which was not the point of this recommendation.



- 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. **[Facebook response (no. 5): “committed to action”]**<sup>31</sup>

---

<sup>31</sup> Facebook’s response only addressed the first sentence of this recommendation and not the specifics that followed. The response addressed how existing transparency reporting practices will disclose data about content moderation during the pandemic. The Board’s recommendation for a specific transparency report on how COVID-19 impacted content moderation, the details this should include that go beyond regular transparency reporting, and the conclusions and lessons learned from this, was not addressed.



## **6. Transparency Report for second quarter of 2021**

This transparency report for the second quarter of 2021 (April 1 – June 30, 2021) sets out key statistics on cases selected by the Board, as well as the decisions and recommendations we made in this quarter.

In the second quarter of 2021, the Board started accepting user appeals to remove content from Facebook and Instagram. We also published five case decisions.



## Key Findings – Second quarter 2021<sup>32</sup>



~207,244 cases submitted to the Oversight Board, of which 9 were submitted by Facebook.

62 user-submitted cases longlisted.

36 user-submitted cases shortlisted.

Nine cases assigned to panel, of which three were submitted by Facebook.

Five cases decided, with 34 recommendations for Facebook.

<sup>32</sup> 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 Q1 but longlisted in Q2 would appear in “submitted cases” in Q1 and “longlisted cases” in Q2. The next footnote provides further context on the numbers of submitted cases.





## Q2 Submitted User Cases

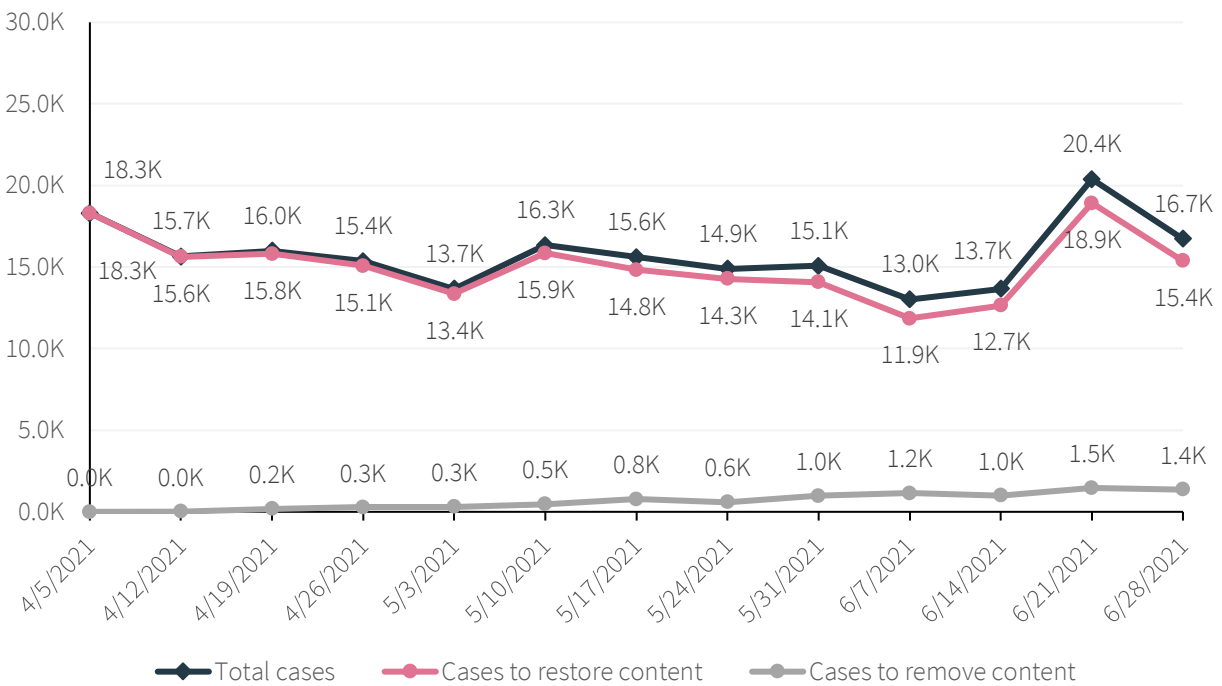
Where users have exhausted Facebook’s appeals process, they can challenge the company’s decision by appealing eligible content to the Oversight Board. On April 13, the Board started accepting user appeals to remove content from Facebook and Instagram.

In this quarter, due to limitations in the functionality of the Case Management Tool (CMT), submitted cases have been counted manually by the Case Selection Team as they were seen in CMT at the time. As such, these numbers should be taken as an estimate and preliminary.

In this period, an estimated **207,235 cases** were submitted by users.

### Estimated number of cases submitted to Oversight Board by week

Number of cases, thousands

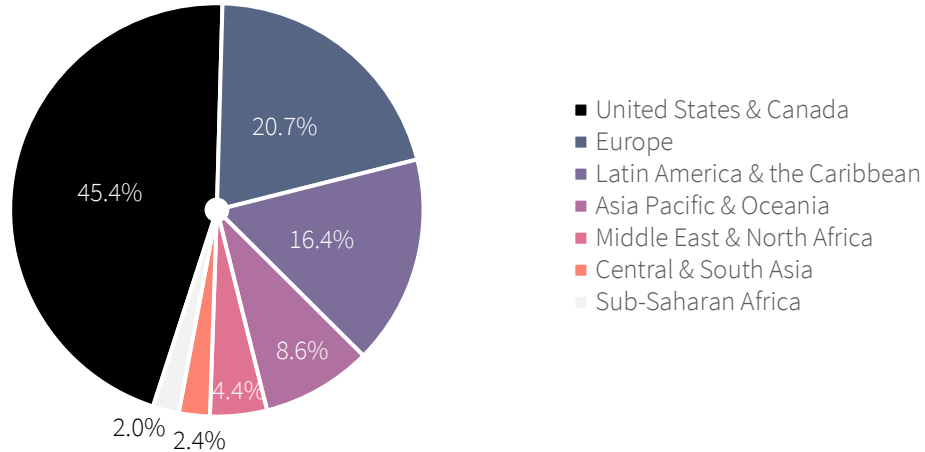




The majority of submitted cases were from the United States & Canada (45.4%) and Europe (20.7%).

### Estimated cases submitted by user-selected region Percent<sup>33</sup>

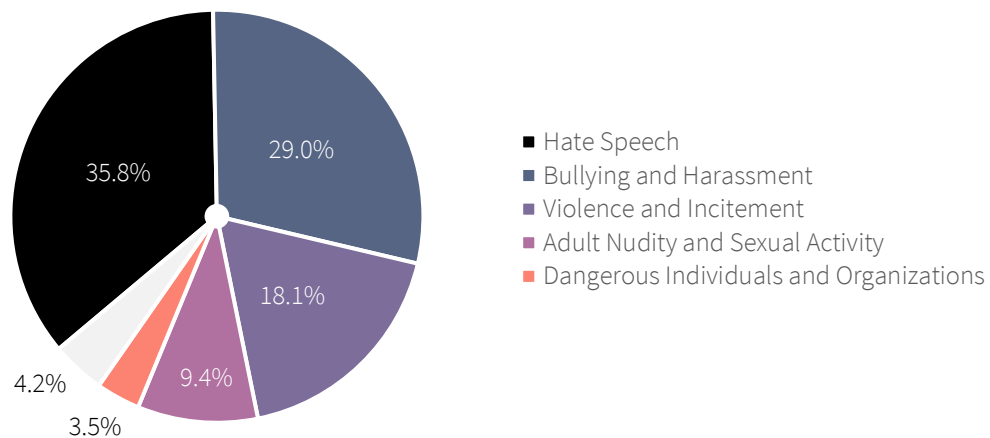
---



Users primarily submitted cases concerning Facebook’s Hate Speech (35.8%) and Bullying and Harassment (29.0%) policies. Other notable Community Standards were Violence and Incitement (18.1%), Adult Nudity and Sexual Activity (9.4%), and Dangerous Individuals and Organizations (3.5%). The chart below only covers cases related to content that has been removed from Facebook and Instagram, and not content which is currently on the platforms as it supposedly does not violate a Community Standard.

### Estimated cases submitted by Community Standard Percent<sup>34</sup>

---



<sup>33</sup> The distribution is approximate and based on a sample of 206,786 cases. Throughout this report, numbers may not add up to 100.0% due to rounding.

<sup>34</sup> The distribution is approximate and based on a sample of 198,510 cases.



Almost all cases submitted by users (98.9%) concerned content shared on Facebook, with only 1.1% of cases concerning content shared on Instagram.

#### Cases submitted by platform<sup>35</sup>

Percent

Facebook	98.9%
Instagram	1.1%

## Q2 Longlisted User Cases

In this period, **62 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 56 pieces of content on Facebook and six pieces of content on Instagram. 43 were cases to restore content and 19 were cases to remove content. The cases involved 60 posts and two comments.

24.2% of the cases are from the Middle East & North Africa and 21.0% were from Europe. Cases were longlisted from 30 countries. The United States, Israel, and Egypt had the highest number of longlisted cases.

#### Longlisted cases by region and country<sup>36</sup>

Number of cases

<b>Middle East &amp; North Africa</b>	<b>15</b>
Algeria	1
Egypt	5
Iran	2
Israel	7
<b>Europe</b>	<b>13</b>
Albania	1
Cyprus	2
France	1
Hungary	1
Italy	2
Spain	2
Sweden	1
Ukraine	2
United Kingdom	1
<b>United States &amp; Canada</b>	<b>9</b>
Canada	1
United States	8

<sup>35</sup> The distribution is approximate and based on a sample of 207,132 cases.

<sup>36</sup> 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.



Sub-Saharan Africa	8
Mauritius	1
Nigeria	4
Somalia	1
South Africa	2
Central & South Asia	7
Bangladesh	1
India	1
Myanmar	4
Pakistan	1
Latin America & the Caribbean	5
Argentina	1
Colombia	1
Mexico	3
Asia-Pacific & Oceania	5
Cambodia	1
China	1
Philippines	2
Vietnam	1
<b>Total</b>	<b>62</b>

About a third of the cases longlisted (35.5%) concerned Facebook’s Hate Speech policy. Another third were requests to remove content (30.6%). Cases regarding Dangerous Individuals and Organizations (14.5%) were also highly represented.

#### Longlisted cases by Community Standard

Number of cases	
Hate Speech	22
Undefined <sup>37</sup>	19
Dangerous Individuals and Organizations	9
Violence and Incitement	3
Bullying and Harassment	3
Adult Nudity and Sexual Activity	3
Sexual Exploitation of Adults	2
Violent and Graphic Content	1
<b>Grand Total</b>	<b>62</b>

<sup>37</sup> 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.



## Q2 Shortlisted User Cases

The Case Selection Committee identifies a shortlist of cases from the longlist to consider for selection. Board Members on the Case Selection Committee rotate every three months, evaluating and selecting cases by majority vote.

In this quarter, the Case Selection Committee reviewed and shortlisted cases on four occasions. While its members generally meet every two weeks, this cadence has been modified when there is no panel with the capacity to accept new cases. In this period, the committee **shortlisted 36 cases**.

The shortlist is sent to Facebook’s legal team to review for legal obligations, as per the Bylaws. Such obligations may mean a case is 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, Facebook has also assessed whether its original decision on a piece of content was correct or not, including deciding to restore content wrongly removed. This does not affect the eligibility of the content under review.

Of the 36 cases shortlisted in this period, 34 were confirmed as eligible by Facebook’s legal team and two cases could not be reviewed by the Board as the content was deleted by the user during the appeals process.

In this quarter, Facebook determined that 12 of the cases shortlisted by the Oversight Board were wrongful applications of Facebook’s Community Standards (See Annex). In all of these cases, Facebook then reversed its original decision and took action on the content.

### Cases where Facebook 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	4	1	5
Dangerous Individuals and Organizations	3	1	4
Undefined <sup>37</sup>	2	0	2
Bullying and Harassment	1	0	1
<b>Total</b>	<b>10</b>	<b>2</b>	<b>12</b>

## Q2 Submitted Facebook Cases

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

Facebook submitted nine cases to the Oversight Board during this period. The Board considered six cases from the previous period and carried over three cases to the next period; in all, 12 Facebook referrals were considered during this period. These covered Facebook’s Child Sexual Exploitation, Abuse and Nudity, Dangerous Individuals and Organizations, Violence and Incitement and Hate Speech Community Standards.



Facebook referrals considered<sup>38</sup>

<i>Case ID</i>	<i>Name</i>	<i>Facebook's Decision</i>	<i>Platform</i>	<i>Language</i>	<i>Community Standard</i>	<i>Countries<sup>39</sup></i>
2021-008-FB-FBR	COVID lockdowns in Brazil	Keep up	FB	Portuguese	N/A	Brazil
N/A	N/A	Keep up	FB	English	N/A	United States
N/A	N/A	Keep up	FB	Dutch	N/A	Netherlands
N/A	N/A	Take down	IG	English	Child Sexual Exploitation, Abuse and Nudity	Italy & United States
N/A	N/A <sup>40</sup>	Keep up	FB	English	N/A	United States
N/A	N/A	Take down	FB	Arabic	Dangerous Individuals and Organizations	Israel
N/A	N/A	Take down	FB	Turkish	Violence and incitement	Turkey
N/A	N/A	Take down	FB	Turkish	Hate Speech	Turkey

<sup>38</sup> Cases which are not selected for assignment do not have a Case ID.

<sup>39</sup> 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.

<sup>40</sup> The original case was deleted by the user after assignment to the panel. Accordingly, Facebook provided a substitute case on a similar topic. The Board decided not to hear the case. For this reason, eleven cases are listed here even though twelve were submitted by Facebook.



N/A	N/A	Keep up	FB	English	N/A	United States
N/A	N/A	Keep up	FB	German	N/A	Bulgaria
N/A	N/A	Keep up	IG	German	N/A	Germany

## Q2 Assigned Cases

The Case Selection Committee assigns cases to panels. In this period, the Case Selection Committee assigned nine cases to panel.

Three cases assigned to a panel were not announced. The COVID origins case was not announced because, soon after assignment, Facebook informed the Board that the page on which the content was posted had accumulated a sufficient number of strikes to be removed from the platform. Accordingly, a decision ordering the restoration of the content by the Board would have been unimplementable by Facebook. Similarly, the misgendering case concerned content on a page that Facebook then removed due to the accumulation of strikes. Facebook later informed the board that the page removal was in error. However, the Board decided not to review the case as the page appeared tailored to content that seemed to violate a Community Standard.

The political candidate case was assigned to panel but was not publicly announced or decided because of concerns raised by Facebook about the content in question relating to the Dangerous Individuals and Organizations Community Standard.

### Cases assigned

<i>Case ID</i>	<i>Name</i>	<i>Date</i>	<i>Platform</i>	<i>Source</i>	<i>Community standard</i>	<i>Countries</i>
2021-006-IG-UA	Ocalan's isolation	4/14/21	IG	User (appeal to restore)	Dangerous Individuals and Organizations	Turkey / United States
2021-007-FB-UA	Myanmar bot	5/13/21	FB	User (appeal to restore)	Hate Speech	Myanmar



N/A	Misgendering	5/13/21	FB	FB	N/A	United States
2021-008- FB-FBR	COVID lockdowns in Brazil	5/26/21	FB	FB	N/A	Brazil
N/A	Political candidate	6/7/21	FB	User (appeal to restore)	Dangerous Individuals and Organizations	Cyprus
N/A	COVID origins	6/10/21	FB	FB	N/A	United States
2021-009- FB-UA	Shared Al Jazeera post	6/23/21	FB	User (appeal to restore)	Dangerous Individuals and Organizations	Israel / Egypt
2021-010- FB-UA	Colombia protests	6/29/21	FB	User (appeal to restore)	Hate Speech	Colombia
2021-011- FB-UA	South Africa slurs	6/30/21	FB	User (appeal to restore)	Hate Speech	South Africa





## Q2 Decided Cases

After being selected, the Board assigns cases to a five-member panel. Members of the panel are randomly chosen but include at least one member from the region implicated in the content and a mix of gender representation.

The panel looks at whether Facebook’s decision is consistent with its content policies and values, and its international human rights responsibilities. The Board’s decisions are binding, and Facebook must implement them within seven days of publication.

In this period, the Oversight Board **decided five cases**. The Oversight Board overturned Facebook’s decision three times and upheld Facebook’s decision twice.

### 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>Countries<sup>41</sup></i>	<i>Outcome</i>
2021-001-FB-FBR	Former President Trump’s suspension	FB	FB	English	Dangerous Individuals and Organizations	United States	Upheld
2021-002-FB-UA	Depiction of Zwarte Piet	FB	User	Dutch	Hate Speech	Netherlands	Upheld
2021-003-FB-UA	Punjabi concern over the RSS in India	FB	User	Punjabi	Dangerous Individuals and Organizations	India	Overturned

<sup>41</sup> 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.



2021-004-FB-UA	Pro-Navalny protests in Russia	FB	User	Russian	Bullying and Harassment	Russia	Overturned
2021-005-FB-UA	“Two buttons” meme	FB	User	English	Cruel and Insensitive / Hate Speech	Armenia, Turkey & United States	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.

#### Human rights standards referenced

Source	2021-001-FB-FBR	2021-002-FB-UA	2021-003-FB-UA	2021-004-FB-UA	2021-005-FB-UA
<b>UN Treaties</b>					
ICCPR <sup>42</sup>					
Non-Discrimination & Remedies (Art. 2)	✓	✓	✓		✓
Life (Art. 6)	✓				
Liberty and security of person (Art. 9)	✓		✓		
Procedural Fairness (Art. 14)					✓
Expression (Art. 19)	✓	✓	✓	✓	✓
Incitement (Art. 20)	✓	✓			
Peaceful Assembly (Art. 21)				✓	
Participation in Public Affairs & Voting (Art. 25)	✓				
Equality (Art. 26)	✓		✓		
ICERD <sup>43</sup>					
Non-Discrimination (Art. 1)	✓				✓
Elimination of Discrimination (Art. 2)		✓			✓
Hate Speech (Art. 4)	✓				
ICESCR <sup>44</sup>					
Non-Discrimination (Art. 2)		✓			
Health (Art. 12)		✓		✓	
Cultural Life (Art. 15)		✓			

<sup>42</sup> International Covenant on Civil and Political Rights

<sup>43</sup> International Convention on the Elimination of All Forms of Racial Discrimination

<sup>44</sup> International Covenant on Economic, Social and Cultural Rights



## Convention on the Rights of the Child (CRC)

Non-Discrimination (Art. 2)	✓
Life, Survival & Development (Art. 6)	✓

## UN Treaty Bodies: Guidance & Recommendations

### Human Rights Committee

General Comment 23 on Minority Rights (CCPR/C/21/Rev.1/Add.5)		✓				
General Comment 29 on States of Emergency (CCPR/C/21/Rev.1/Add.11)		✓				
General Comment 31 on General Legal Obligation (CCPR/C/21/Rev.1/Add. 13)	✓		✓			
General Comment 32 on Right to Equality Before Courts & Fair Trial (CCPR/C/GC/32)						✓
General Comment 34 on the Freedoms of Opinion and Expression (CCPR/C/GC/34)	✓	✓	✓	✓	✓	✓
General Comment 35 on the Right to Liberty & Security of Person (CCPR/C/GC/35)				✓		

### Committee on Economic, Social & Cultural Rights

General Comment 21 on the Right to Take Part in Cultural Life (E/C.12/GC/21)		✓				
--	--	---	--	--	--	--

### Committee on the Elimination of Racial Discrimination

General Recommendation 34 on People of African Descent (CERD/C/GC/34)		✓				
Concluding Observations on the Netherlands (CERD/C/NLD/CO/19-21)		✓				

## Other UN Human Rights Standards

### Responsibilities of Businesses

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

### Incitement

Rabat Plan of Action (A/HRC/22/17/Add.4)	✓	✓			✓
--	---	---	--	--	---

### UN SR Freedom of Expression Reports

Online Hate Speech (A/74/486)		✓	✓	✓	✓
Online Content Regulation (A/HRC/38/35)	✓	✓	✓		✓
Artistic Expression (A/HRC/44/49/Add.2)					✓
Mission to Turkey (A/HRC/35/22/Add.3)					✓
Follow-Up on Country Visits (A/HRC/41/35/Add.2)					✓

### Joint Declarations and Statements of UN & Regional Freedom of Expression Mandates

Promotion of Freedom of Expression during Pandemic	✓				
"Fake News," Disinformation & Propaganda					✓



Racism	
UN Special Rapporteur on Racism, Visit to the Netherlands (A/HRC/44/57/Add.2)	✓
UN Working Group of Experts on People of African Descent, Visit to the Netherlands (A/HRC/30/56/Add.1)	✓
UN SR Minority Issues Reports	
Interpretation of the Declaration on the Rights of Minorities (A/HRC/22/49)	✓
Social Media & Hate Speech (A/HRC/46/57)	✓
UN SR Violence Against Women Reports	
Online Violence and Violence Facilitated by ICT (A/HRC/38/47)	✓
UN SR Peaceful Assembly and Association Reports	
Best practices that promote and protect the rights to freedom of peaceful assembly and of association (A/HRC/20/27)	✓

### Decision timelines

Except for exceptional circumstances, the Bylaws specify that case decisions and implementation will occur a maximum of 90 days from the date the case is selected for review by the Case Selection Committee. For the six decisions the Oversight Board published in this quarter, the average number of days from assignment of case to implementation of the Board’s decision has been 88 days.

<i>Case ID</i>	<i>Name</i>	<i>Beginning of 90-day period</i>	<i>Board’s decision published</i>	<i>Facebook implements decision</i>	<i>Number of days taken out of 90 days</i>
2021-001-FB-FBR	Former President Trump’s suspension	1/21/21	5/5/21	N/A (Upheld FB’s decision)	104 <sup>45</sup>
2021-002-FB-UA	Depiction of Zwarte Piet	1/26/21	4/13/21	N/A (Upheld FB’s decision)	77

<sup>45</sup> In this case, the Board’s commitment to carefully reviewing the more than 9,000 public comments submitted extended the case timeline beyond 90 days, in line with the Board’s Bylaws.



2021-003-FB-UA	Punjabi concern over the RSS in India	2/5/21	4/29/21	N/A (FB already restored content)	83
----------------	---------------------------------------	--------	---------	-----------------------------------	----

2021-004-FB-UA	Pro-Navalny protests in Russia	2/25/21	5/26/21	5/26/21	90
----------------	--------------------------------	---------	---------	---------	----

2021-005-FB-UA	"Two buttons" meme	2/25/21	5/20/21	5/20/21	84
----------------	--------------------	---------	---------	---------	----

### Questions for Facebook

To assist with making its decisions, the Oversight Board sends questions to Facebook. Of the 113 questions sent by the Oversight Board to Facebook about decisions published in this quarter, Facebook answered 92 questions, partially answered 12 questions and declined to answer nine.

Information on the questions which Facebook did not answer on the case related to former President Trump can be found in the [Board's decision](#) (Section 6, final paragraph).

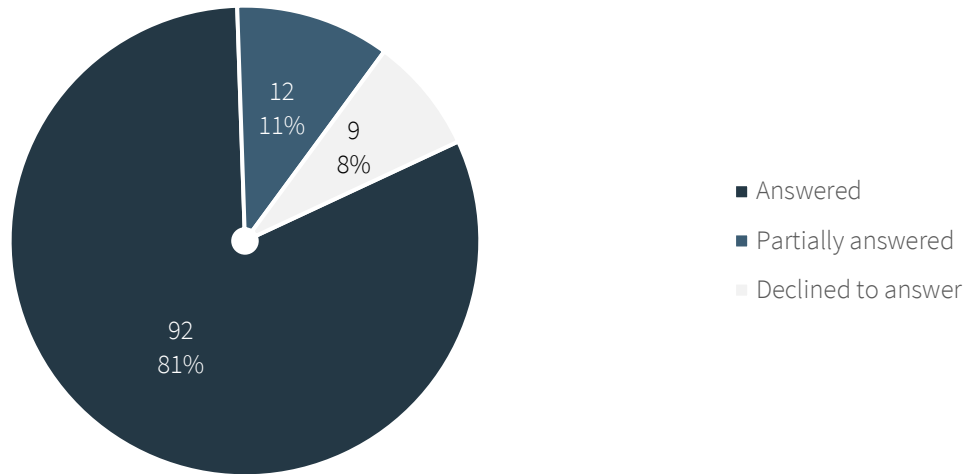
### Oversight Board questions answered by Facebook

Number of questions

<i>Case ID</i>	<i>Name</i>	<i>Answered</i>	<i>Partially answered</i>	<i>Declined to answer</i>	<i>Total</i>
2021-001-FB-FBR	Former President Trump's suspension	37	2	7	46
2021-002-FB-UA	Depiction of Zwarte Piet	12	0	0	12
2021-003-FB-UA	Punjabi concern over the RSS in India	13	5	2	20
2021-004-FB-UA	Pro-Navalny protests in Russia	18	2	0	20
2021-005-FB-UA	"Two buttons" meme	12	3	0	15
<b>Total</b>		<b>92</b>	<b>12</b>	<b>9</b>	<b>113</b>



## Oversight Board questions answered by Facebook



### Public Comments

The Oversight Board conducts a public comment process to assist it in its decision making. In this period, the Oversight Board received 9,746 comments, 7,726 of which were published.

### Public comments received by publication status

Number of comments

Case ID	Name	Comments published	Comments not published (no consent)	Comments not published (violated terms)	Total	Comments unattributed <sup>46</sup>
2021-001-FB-FBR	Former President Trump's suspension	7,656	1,447	563	9,666	2,589
2021-002-FB-UA	Depiction of Zwarte Piet	22	0	0	22	8
2021-003-FB-UA	Punjabi concern over the RSS in India	6	1	5	12	4
2021-004-FB-UA	Pro-Navalny protests in Russia	23	0	0	23	4
2021-005-FB-UA	"Two buttons" meme	19	3	1	23	11
<b>Total</b>		<b>7,726</b>	<b>1,451</b>	<b>569</b>	<b>9,746</b>	<b>2,616</b>

<sup>46</sup> Unattributed comments are published comments with the author's name redacted by request.



The majority of public comments (96.1%) came from individuals, while a minority came from organizations (3.9%).

### 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>
2021-001-FB-FBR	Former President Trump's suspension	9,299	367	9,666
2021-002-FB-UA	Depiction of Zwarte Piet	18	4	22
2021-003-FB-UA	Punjabi concern over the RSS in India	9	3	12
2021-004-FB-UA	Pro-Navalny protests in Russia	18	5	23
2021-005-FB-UA	"Two buttons" meme	21	2	23
<b>Total</b>		<b>9,365</b>	<b>381</b>	<b>9,746</b>

The vast majority of public comments (96.9%) in this quarter came from the US and Canada.

### Public comments received by region

Number of comments

<i>Region</i>	<i>2021-001-FB-FBR</i>	<i>2021-002-FB-UA</i>	<i>2021-003-FB-UA</i>	<i>2021-004-FB-UA</i>	<i>2021-005-FB-UA</i>	<i>Total</i>
United States & Canada	9,388	15	7	18	13	9,441
Europe	136	7	4	4	8	159
Asia Pacific & Oceania	80	0	0	0	1	81
Latin America & Caribbean	23	0	1	0	1	25
Sub-Saharan Africa	19	0	0	0	0	19
Middle East & North Africa	13	0	0	1	0	14
Central & South Asia	7	0	0	0	0	7
<b>Total</b>	<b>9,666</b>	<b>22</b>	<b>12</b>	<b>23</b>	<b>23</b>	<b>9,746</b>



## Recommendations

In addition to providing decisions on appealed content, the Oversight Board also provided 34 policy recommendations to Facebook, which the company responded to publicly within 30 days.

These included recommendations on content policy (clarification or changes to rules), enforcement (clarification or changes to how rules are applied), and transparency (on disclosure of information to the public).

In Q2 2021, the Board’s Implementation and Working Group was established and met two times. This brings together a group of Board Members to review Facebook’s implementation of case decisions and policy recommendations. Facebook representatives are invited to these meetings to answer the Board’s questions about their responses to recommendations and steps being taken to act on commitments they have made.

Below, the Board lists all policy recommendations listed in case decisions during Q2. Facebook has not always directly responded to the Board’s recommendations in the same format that the Board set out in case decisions. Therefore, the Board has made its best effort to map Facebook’s commitments to the Board’s policy guidance as originally stated in each case, which sometimes means there are a different total number of recommendations per decision. The Board has reproduced Facebook’s initial commitments made within the 30-day response window. Facebook continues to update their on-going progress to implement the recommendations. The Board will continue to monitor Facebook’s implementation efforts and include information about this monitoring in future transparency reporting.

### Oversight Board recommendations to Facebook

Number of recommendations

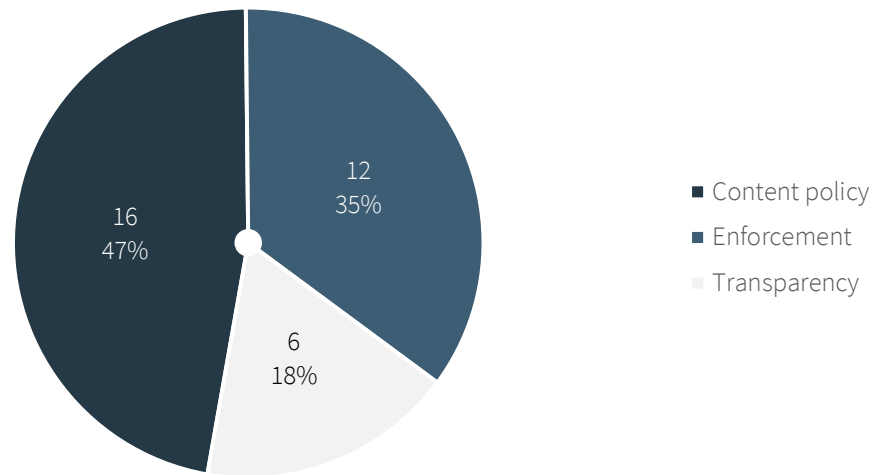
<i>Case ID</i>	<i>Name</i>	<i>Content policy</i>	<i>Enforcement</i>	<i>Transparency</i>	<i>Total</i>
2021-001-FB-FBR	Former President Trump’s	8	5	5	18
2021-002-FB-UA	Depiction of Zwarte Piet	1	1	0	2
2021-003-FB-UA	Punjabi concern over the RSS in India	1	1	1	3
2021-004-FB-UA	Pro-Navalny protests in Russia	5	1	0	6
2021-005-FB-UA	“Two buttons” meme	1	4	0	5
<b>Total</b>		<b>16</b>	<b>12</b>	<b>6</b>	<b>34</b>





## Oversight Board recommendations to Facebook

---



### Content policy recommendations

Recommendations to clarify rules, for rules to be modified, or for new rules to be created.

#### [Former President Trump's suspension \(2021-001-FB-FBR\)](#)

- 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. **[Facebook responses (no. 2 and no. 3): “implementing fully”<sup>47</sup>]**
- 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

---

<sup>47</sup> Facebook split this recommendation in two for its response. The Board notes that Facebook did not include the first three sentences here when framing their response, though the Board considers this part of the guidance it provided on this issue. The Board notes that Facebook's response does not address the use of the Rabat Plan of Action factors in incitement cases.



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. **[Facebook response (no. 4): “implementing fully”]**

- Periods of suspension should be long enough to deter misconduct and may, in appropriate cases, include account or page deletion. **[Facebook response (no. 5): “implementing fully”]**
- 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. **[Facebook response (no. 8 and no.9): “implementing fully”<sup>48</sup>]**
- 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. **[Facebook response (no. 11): “implementing fully”]**
- 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. **[Facebook response (no. 12): “implementing fully”<sup>49</sup>]**
- 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. **[Facebook response (no. 16): “implementing fully”<sup>50</sup>]**

---

<sup>48</sup> Facebook’s response as framed under No. 8 departs from the Board’s guidance: (i) the new policy defines periods for account suspension that can apply *only to* “influential users” *and only* “during times of civil unrest and ongoing violence” (i.e. suspension periods of more than a month are only available when particular circumstances are present and the user has a particular status, whereas at other times or when the user is not influential, permanent deletion is the only available option), and (ii) Facebook’s response focuses on risk to public safety broadly at the end of the suspension period, rather than specific risks from the suspended user.

<sup>49</sup> Facebook’s response, and the article it links to in the transparency center, does not address the criteria for accounts or pages being included in the “cross-check” system (the Board understands the cross-check system does not apply to all high-visibility content, but is a process that applies to specific accounts or pages based on their identity).

<sup>50</sup> Facebook’s response indicates this information has been added to the Transparency Center but does not indicate whether or when this information will be added to, and/or linked from, the Community Standards and Community Guidelines. The response and linked article does not address if policy allowances or exceptions can be applied to strikes and penalties.



- 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. [Facebook response (no. 19): “implementing fully”]

#### [Depiction of Zwarte Piet \(2021-002-FB-UA\)](#)

- 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. [Facebook response (no. 1): “committed to action”]

#### [Punjabi concern over the RSS in India \(2021-003-FB-UA\)](#)

- 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. [Facebook response (no. 1): “committed to action”]

#### [Pro-Navalny protests in Russia \(2021-004-FB-UA\)](#)

- Explain the relationship between the policy rationale and the “Do not’s” as well as the other rules restricting content that follow it. [Facebook response (no. 1): “implementing in part”]
- 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. [Facebook response (no. 2): “assessing feasibility”]
- 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. [Facebook response (no. 3): “implementing fully”]
- 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. [Facebook response (no. 4): “implementing in part”]
- 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. [Facebook response (no. 5): “assessing feasibility”]

#### [“Two buttons” meme \(2021-005-FB-UA\)](#)

- Include the satire exception, which is currently not communicated to users, in the public language of the Hate Speech Community Standard. [Facebook response (no. 2): “implementing fully”]



## Enforcement recommendations

Recommendations to change content moderation enforcement processes, including about the use of automated or human review, notification of enforcement action to users, and access to appeals.

### [Former President Trump’s suspension \(2021-001-FB-FBR\)](#)

- 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. **[Facebook response (no. 1): “implementing fully”]**
- 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. **[Facebook response (no. 6): “implementing fully”]**
- 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. **[Facebook response (no. 7): “implementing fully”<sup>51</sup>]**
- 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. **[No response]**
- 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. **[Facebook response (no. 17): “implementing fully”]**

### [Depiction of Zwarte Piet \(2021-002-FB-UA\)](#)

- 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. **[Facebook response (no. 2): “committed to action”]**

### [Punjabi concern over the RSS in India \(2021-003-FB-UA\)](#)

- 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

---

<sup>51</sup> The language of the policy recommendation was reframed in Facebook’s response, and does not address the Board’s concern regarding undue influence.



possible while fully protecting the health of Facebook’s staff and contractors. **Facebook response (no. 2): “committed to action”**

[Pro-Navalny protests in Russia \(2021-004-FB-UA\)](#)

- 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. **[Facebook response (no. 6): “assessing feasibility”]**

[“Two buttons” meme \(2021-005-FB-UA\)](#)

- 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. **[Facebook response (no. 1): “assessing feasibility”]**
- Make sure that Facebook 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. **[Facebook response (no. 3): “assessing feasibility”]**<sup>52</sup>
- 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. **[Facebook response (no. 4): “assessing feasibility”]**
- Ensure appeals based on policy exceptions are prioritized for human review. **[Facebook response (no. 5): “assessing feasibility”]**

## Transparency

Recommendations to increase the transparency of Facebook’s content moderation, including through its transparency reports.

[Former President Trump’s suspension \(2021-001-FB-FBR\)](#)

- 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. **[Facebook response (no. 10): “implementing fully”]**
- Facebook should report on the relative error rates and thematic consistency of determinations made through the cross-check process compared with ordinary enforcement procedures. **[Facebook response (no. 13): “no further action”]**

---

<sup>52</sup> The Board notes that Facebook summarized this recommendation when formulating their response.



- 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. [Facebook response (no. 14): “implementing in part”<sup>53</sup>]
- 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. [Facebook response (no. 15): “assessing feasibility”]
- 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. [Facebook response (no. 18): “assessing feasibility”]

#### [Punjabi concern over the RSS in India \(2021-003-FB-UA\)](#)

- 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. [Facebook response (no. 3): “assessing feasibility.”]

---

<sup>53</sup> Facebook’s response focused on the first sentence of the Board’s guidance, and did not address the importance of an open reflection on design and policy choices.



## 7. Annex - Summaries of cases where Facebook identified its original decision on a piece of content was incorrect

### Q1 2021

1. Even though Facebook identified this case as an enforcement error, it was assigned to panel by the Oversight Board as 2021-006-IG-UA (“Ocalan’s isolation”).
2. The content is an editorial by a Pakistani news channel criticizing the Pakistani government's approach in dealing with Khadim Hussain Rizvi and his party, Tehreek-e-Labbaik Pakistan (TLP). The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee shortlisted this case, and it was therefore referred to Facebook for a legal-eligibility review. Facebook restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that although the individual and organization could fall under Facebook's list of Dangerous Individuals and Organizations, the content is critical of Rizvi and the TLP, without any praise, support or representation as required under the Community Standards.<sup>54</sup> The case illustrates the importance of context in determining whether references to dangerous organizations or individuals constitute praise, condemnation, or mere factual reporting.
3. The content is a screenshot of a tweet about Shamima Begum who left the UK to join ISIS and was groomed as a young teen (subsequently losing her citizenship to the UK as a result in an order decreed by the Supreme Court). The tweet says that her situation should garner pity, not scorn. The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that although Begum has been a member of ISIS and thus could fall under Facebook’s list of Dangerous Individuals and Organizations, the content does not praise, support or represent Begum or ISIS as required under the Community Standards.
4. The content contains a photo of two men kissing. It also contains text with the hashtag #proudboys, followed by references to LGBTQ+ pride and condemnation of violence against peaceful protesters. The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that although the Proud Boys could fall under Facebook's list of

---

<sup>54</sup> In each of the short summaries in this annex, the sentence which includes the words “it seems that” represents an assessment by Oversight Board Administration staff of why Facebook may have reversed its original decision on this content, which may differ from Facebook’s actual reason for reversing its decision.



Dangerous Individuals and Organizations, the content does not praise, support or represent the Proud Boys but instead alluded to the organization as a form of satire. The case illustrates the importance of context when satire or sarcasm is involved.

5. The content is a post calling attention to ethnic Oromo political prisoners who are on hunger strike in Ethiopia. It demands their release, calls for unity of Oromo political parties and pushes for the demolition of the "Neftegna" (referring to the Amhara people) system of government. The post also contains a screenshot of a post from another user depicting an Oromo politician as a monkey. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content was restored possibly because "Neftegna" is ambiguous in meaning and might not be considered by Facebook as rising to the level of a slur, and the comparison between a politician and a monkey might not implicate any protected characteristics.
6. The content is a screenshot regarding the Indian farmer protests that claims the famous singer Rihanna is Muslim, and that she became famous so she could destabilize India by supporting the farmers' protests. The user has written a caption above the screenshot that appears to condemn it, and compares these false statements about Rihanna to white supremacy misinformation. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that nothing in the content would violate the Community Standard on Hate Speech.
7. The content was posted by a user to draw attention to the death of their neighbor, allegedly in a murder by the neighbor's husband. The post further calls out violence against women in Israel as a deep-rooted cultural issue. It was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that while the content could be interpreted as calling for the murder of women, in fact the user is likely to have meant to question violence against women.
8. The content is a video showing some visually disturbing parts of the pig slaughtering process. The post also contains ironic text asking if anyone would be interested to be paid for performing a list of duties which appear to be mistreatment of animals. In the user statement, the user explains that the post intended to raise awareness on the cruelties behind meat production. The content was erroneously taken down under the Community Standard on Coordinating Harm and Publicizing Crime. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content is sarcastic and, despite depicting mistreatment of animals, does not constitute statements of intent or calls to action representing, supporting, advocating for, promoting, or admitting to these acts.





9. The content is a post that seems to be an adaptation of the original play “F\*ggot” (“M\*rica” in Spanish) by an Argentinian writer, and attempts to reclaim the use of the slur “m\*rica” and the word “puto” by repeating them and casting them in a sarcastic light. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content was removed without taking into account the full context or user intent.
10. The content is a post in Portuguese that contains (,) in the middle of a phrase which, when translated correctly in our view, reads “Don't hit women, it's a crime!”. However, when translated by Facebook’s automated tools, the comma is removed, changing the phrase to “Hitting women is not a crime!”. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content was possibly removed by automation, which presented issues with detection and translation accuracy.
11. Posted by a user from the US around January 7, 2021, the content states that all riots matter, seeming to draw a comparison between pro-Trump protestors who sought to overturn the presidential election at the U.S. Capitol on January 6, 2021, and protestors in the Black Lives Matter movement. The content suggests that supporters of both protests condemn destructive riots only when the riots are contrary to their viewpoints. The content was erroneously taken down under the Community Standard on Coordinating Harm and Publicizing Crime. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content could be interpreted as being critical of both events and does not contain statements of intent or calls to action representing, supporting, advocating for, promoting, depicting, or admitting to these events.
12. The content is a post about a university in Ethiopia erecting a statue of Hachalu Hundessa, an Ethiopian Oromo pop star and civil rights activist who was assassinated in June 2020 and whose death sparked protests that led to 160 deaths. The post accuses the “Neftegna” (referring to the Amhara people) of sitting on the statue and asking for the horse in the statue to be changed. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that content was restored possibly because “Neftegna” is ambiguous in meaning and might not be considered by Facebook as rising to the level of a slur.
13. The content contains a video depicting two scenes of protestors trespassing in the US Capitol Building on January 6, 2021. Various groups of protestors can be seen casually walking through hallways and up the stairs, accompanied by mostly unclear shouting. The accompanying caption uses neutral language and labels the video as breaking news. The content was erroneously taken



down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that it is uncertain whether Facebook has classified any of the participants or organizations present at the Capitol on January 6, 2021, as Dangerous Individuals and Organizations, but believes that even if it has, the content is neutral and does not praise, support or represent such individuals or organizations.

14. The content was posted on a journalism organization's page and is a screenshot of the headline in a news article, which is seemingly a threat from the Bajrang Dal (a Hindu nationalist organization) to beat Hindus who visit churches on Christmas Day in India. The content condemns "extremist groups like the Bajrang Dal" which are "against the idea of India". The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that while the headline from the article, "Hindus will get brutally beaten", seems to call for violence against the protected characteristic of ethnicity or religion, the content itself condemns the group making this threat.
15. In an India-based, limited-audience Facebook group, one user uploaded a screenshot of a photograph on Instagram showing a man at farmers' protests, dressed in military fatigues, holding a sign which says that if, as some claim, the protesting farmers are terrorists then he is one also. The content at question is a comment on this post asserting that the Indian military are in fact terrorists, adding a shrugging emoji. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the Indian military refers to a governmental body instead of the protected characteristic of national origin.
16. Even though Facebook identified this case as an enforcement error, it was assigned to panel by the Oversight Board and published as case decision 2021-003-FB-UA ("Punjabi concern over the RSS in India").
17. The content is an image of Ralph Wiggum, a character from the American TV show 'The Simpsons.' He is depicted saying "I am a Proud Boy!" and apparently urinating in his pants. The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that although the Proud Boys could fall under Facebook's list of Dangerous Individuals and Organizations, the content does not praise, support, or represent the Proud Boys but instead alluded to the organization as a form of satire.



## Q2 2021

1. Even though Facebook identified this case as an enforcement error, it was assigned to panel by the Oversight Board as 2021-009-FB-UA (“Shared Al Jazeera post”).
2. The content is a user's sharing of a post by TRT news which includes excerpts from a statement made by al-Qassam Brigades, saying that they had fired at Ashdod and Ashkelon in response to the targeting of "safe homes and resistance men", and that "what is coming is greater". The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the takedown may present similar issues as those in 2021-009-FB-UA in handling commentary on the newsworthy activities of organizations possibly designated as dangerous.
3. The content is a meme depicting the occupation of Palestine in a cartoon image that shows a boy (next to a Palestinian flag icon) sleeping with his dog (next to an Israeli flag icon). The dog is initially curled up on the floor but slowly starts to dominate the bed as time goes by. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the Israeli flag is possibly taken to represent the country or government of Israel instead of people based on the protected characteristic of nationality and that the comparison may not be dehumanizing.
4. The content is a meme which is part of an Instagram gallery. In the meme, two women exchange ideas about the Palestinian-Israeli conflict. One woman says that Israelis are oppressors and Palestinians are oppressed, and that the situation is about everything but religion. The content was erroneously taken down under the Community Standard on Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that characterization as oppressors may not fall under any category of Hate Speech in the Community Standards.
5. The content is a post containing two photos of Boko Haram leader Abubakar Shekau, one of them depicting him with an ISIS flag. The caption to the photos says that the user prays this would be his last and ever death. In the user statement the user explains that this was posted to celebrate Shekau's death. The content was erroneously taken down under the Community Standard on Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the user, by celebrating Shekau's death, is effectively condemning him.
6. The content is a post in Burmese showing one page of a document that seems to provide of list of two people deemed affiliated with/supportive of the military dictatorship in Myanmar and was posted as an act of protest. The content was erroneously taken down under the Community Standard of Bullying and Harassment. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content did not rise to the level of "calling for, or making



statements of intent to engage in, bullying and/or harassment,” of individuals who may be public figures, or that the content may have been deemed legitimate political speech in light of heightened tensions in the region.

7. The content is a video portraying a discussion between two individuals where it seems one is likely a gay man; an insertion of another short clip features a man saying that the cure for such individuals is to beat them, which is then followed by another clip of the gay individual being attacked. The content was erroneously kept up under the platform. The Case Selection Committee referred this appeal to Facebook, who removed the content under its Violent and Graphic Content Community Standard. Without explanation from Facebook on either the keep up or the removal, it seems that the content was finally removed as it appeared to endorse and depict violence against the LGBTQ+ community.
8. The content is a post quoting Mohammad Bagher Ghalibaf, Speaker of the Parliament of Iran, for his reaction to negative comments on Qassem Soleimani and engaging in debate on how the prominent leader's role in history should be remembered. The content was erroneously taken down under the Community Standard of Dangerous Individuals and Organizations. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the content was posted for the purposes of sharing news or information and did not praise the designated individual.
9. The content is a post saying that the user does not care what "illegals" experience at the American border and that they should not be there. The content was erroneously taken down under the Community Standard of Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that, on the assumption that the term "illegals" refers to undocumented migrants and implicates the quasi-protected characteristic of immigration status, this term might not rise to the level of "[d]ehumanising speech or imagery in the form of comparisons, generalisations...to or about...[o]ther criminals"; it might not be a call for exclusion to say that they should not be there and even if it was, calls for exclusion would not be Hate Speech for a quasi-protected characteristic.
10. The content is a post discussing the nickname of the South African national football team, "bafana bafana," a Zulu phrase meaning "the boys, the boys" or "go boys, go boys" which the user finds demeaning for a team of adult men, but which is widely regarded as affectionate. The content was erroneously taken down under the Community Standard of Hate Speech. The Case Selection Committee referred this appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that the phrase might not rise to the level of a slur or otherwise violate the Community Standard of Hate Speech.
11. The content is a post that discusses at length events in 1913 where Cham Albanians were killed by Greek armed forces. It brings up the forced migration of the Cham people as well as current-day issues of repatriation and the acknowledgement of history. The post was erroneously taken down under the Community Standard of Hate Speech. The Case Selection Committee referred this



appeal to Facebook, who restored the content. Without explanation from Facebook on either the removal or the restoration, it seems that although the mention of "Greek barbaric Deli Janaqi" in the post associated "Greek" with "barbaric", the term "barbaric" is directed at an individual (Deli Janaqi), instead of the protected characteristic of ethnicity or national origin (the Greeks).

12. The content is a post by a Philippines official accusing certain people opposing the government of affiliations with the communist rebellion when such affiliations were unverified. A user reported the content, arguing that this practice of "redtagging" could promote hate crimes and violence against the people identified. The content was erroneously kept up under the platform. The Case Selection Committee referred this appeal to Facebook, who removed the content under its Community Standard on Violence and Incitement. Without explanation from Facebook on either the keep up or the removal, it seems that the content possibly "indicates knowledge of or shares sensitive information that could expose others to harm" under the Community Standard. The content could also raise issues on privacy, misinformation, and protection for alleged "communists" in the Philippines.