March 2022

Meta Q4 2021 Quarterly Update on the Oversight Board

Meta

TABLE OF CONTENTS

Introduction **3**

I. Meta Content Referrals **4**

II. Progress on Non-Binding Recommendations **6**

How to Read This Update 9

Update on Non-Binding Recommendations 11

Transparency 11

Policy 12

Enforcement 13

III. Appendix **15**

How to Read This Appendix 15

Appendix A. Transparency 16

Appendix B. Policy 24

Appendix C. Enforcement 33

IV. Index **36**

Introduction

We are committed to publishing regular updates to give our community visibility into our responses to the Oversight Board's independent decisions about some of the most difficult content decisions Meta makes. These quarterly updates provide regular check-ins on the progress of this long-term work and share more about how Meta approaches decisions and recommendations from the board. This update, covering decisions the board issued in Q4 2021, includes sections that detail (1) our content decision referrals and Policy Advisory Opinion (PAO) requests to the board and (2) our progress on implementing the board's non-binding recommendations. The report is meant to hold us accountable to the board and the public.

I. Meta Content Referrals

In addition to providing people who use Facebook and Instagram with direct access to appeal to the board, we regularly and proactively identify some of the most significant and difficult content decisions and ask the board to review them. We previously outlined how we prioritize cases we believe are significant and difficult for Meta content referrals in our Newsroom. The questions at issue in these referrals generally involve real-world impact and issues that are severe, large-scale, and/or important for public discourse. Additionally, the referrals of content and Policy Advisory Opinion (PAO) requests raise questions about current policies and their enforcement.

For Meta content referrals, the process begins with an internal review of content decisions that are geographically diverse, cover questions about a wide range of policies found in our Facebook Community Standards or Instagram Community Guidelines, and represent both content removed as well as left up. Then, teams with expertise on our content policies, our enforcement processes, and specific cultural nuances from regions around the world review the candidate cases and provide feedback on both their significance and difficulty. At the end of this process, we refer the most significant and difficult content decisions to the board. The board has sole discretion to agree or decline to review the decisions referred through this process. As with user appeals about Meta's content decisions, the board's decision on Meta content referrals is binding.

For PAO requests, we ask the board to advise us on our policies and content moderation systems more generally. Once the board issues the PAO, we will consider and publicly respond to its recommendations within 60 days.[1] While not binding, the board's guidance through PAOs is crucial to our ongoing commitment for the board to hold Meta accountable for our policies, processes, and decisions.

Between October 1, 2021 and December 31, 2021, we referred 4 content cases and 1 PAO request to the board, and the board selected 1 content case and 1 PAO request[2]:

1. A case about a post detailing sexual violence against minors in Sweden [link]
2. A PAO regarding our cross-check system [link].

We will continue to refer content decision cases to the Oversight Board based on the process described above.

[1] To facilitate a more comprehensive response to board recommendations, on February 1st, 2022 the board's bylaws were updated to reflect that Meta now has 60 days instead of 30 days to respond.
[2] Per the bylaws, the board has 90 days to select a Meta-referred case. These are the numbers as of this report's publication.

Q4 Meta referred content decision case breakdown

● TOTAL Q4 ● TOTAL THROUGH Q4 2021 (12/31/21)

5	49

Total number of referrals
(Meta content referrals + policy advisory opinions)

1	2

Total number of policy advisory opinion requests

1	2

Total number of policy advisory opinion requests selected

4	47

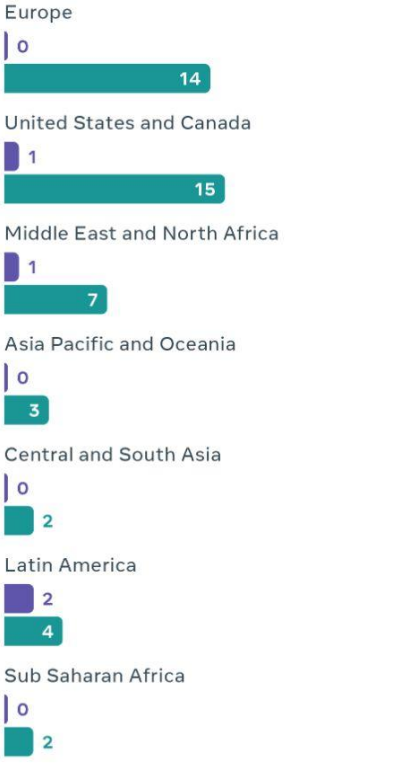
Total number of Meta content referrals[1]

1	5

Total number of Meta content referrals selected

POLICY VIOLATION

REGION[3]



[1] The numbers in the Policy Violation and Region sections of this table only apply to Meta Content Referrals, not Policy Advisory Opinion Requests.

[2] When we decide to leave content up, there is, by definition, no policy violation. As a result, we categorize the policy violation as "not applicable" for referrals of content we left up on Facebook and Instagram.

[3] Meta defines "region" according to an analysis of several factors, including the location of the posting user, the language(s) the content includes, and countries/regions referenced in the content.

II. Progress on Non-Binding Recommendations

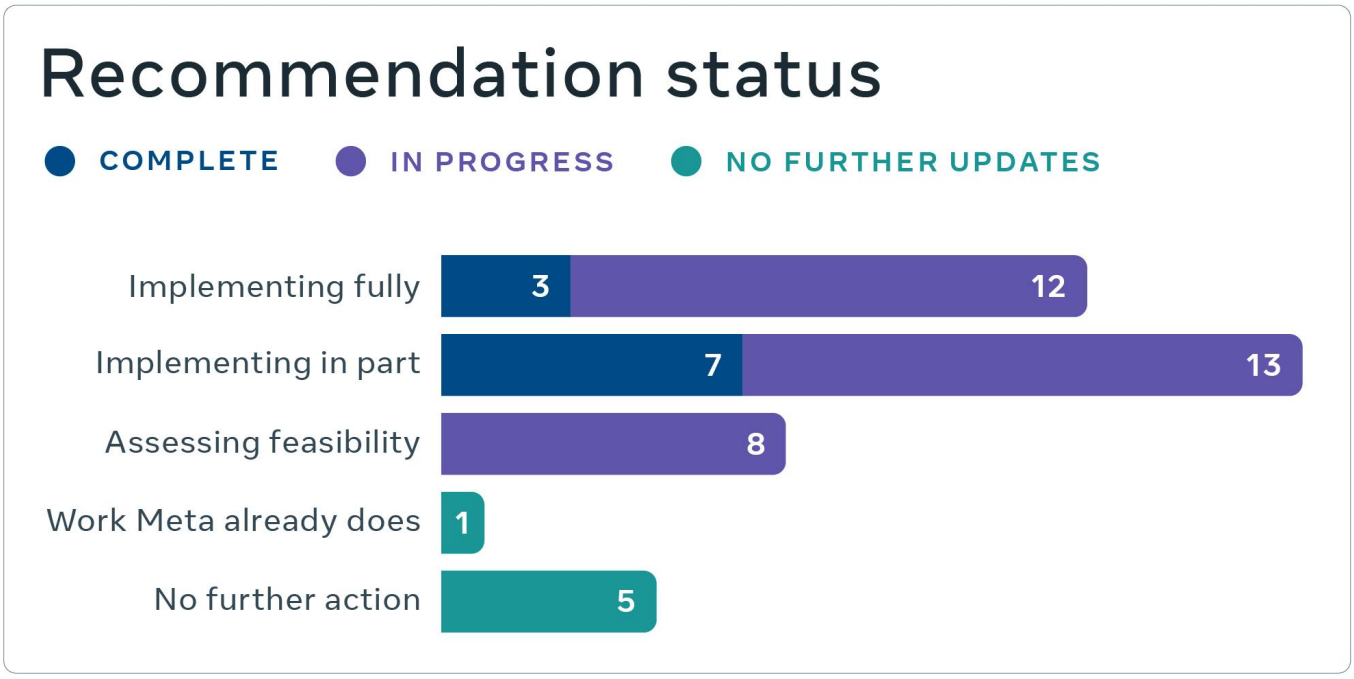
This section provides a detailed update on how we continue to address the board's non-binding recommendations. From January through December 2021, the board issued 87 non-binding recommendations. In Q4 2021, the board issued 9 recommendations. In this update, we address those 9 new recommendations and 40 others that carried over from our Q2 and Q3 2021 Quarterly Update on the Oversight Board.

Below is a graph depicting the status of each of the 49 recommendations. The categorization of our responses to the board's recommendations include the following:

- **Implementing fully:** We agree with the recommendation and have or will implement it in full.
- **Implementing in part:** We agree with the overall aim of the recommendation and have or will implement work related to the board's guidance.
- **Assessing feasibility:** We are assessing the feasibility and impact of the recommendation.
- **No further action:** We will not implement the recommendation, for example, due to a lack of feasibility or disagreement about how to reach the desired outcome.
- **Work Meta already does:** We have addressed the recommendation through an action that we already do.

The current status for our responses to the board's recommendations include the following:

- **Complete**: We have completed full or partial implementation in line with our response to the board's recommendation, and will have no further updates on the recommendation.
- **In progress:** We are continuing to make progress on our response to the board's recommendation, and will have further updates on the recommendation.
- **No further updates:** We will not implement the recommendation or have addressed the recommendation through an action that we already do, and will have no further updates on the recommendation.



As this update highlights, the impact of the board continues to go beyond its binding decisions to uphold or overturn our enforcement decision on a single piece of content. In particular, this quarter we saw the impact of previous recommendations becoming part of Meta's systems and policies. Through its recommendations, the board has pushed us to be more transparent about how our systems operate, how we evaluate trade-offs in content moderation enforcement, and how we write our policies. These recommendations have had broad impact, and have challenged us to think differently about how we can make changes to our content moderation policies and enforcement.

For example, in Q4 2021, because of the board's recommendations we:

- **Completed analyses of proposals for modifying our user notifications, including:**
  - Running a test to measure the effects of informing people if automation or manual review resulted in their content being removed. Based on our findings, we are proceeding with plans to roll out this new user messaging in the first half of this year.
  - Exploring how frequently a violation changes on appeal.

- **Completed four updates to the Community Standards:**
  - Updated our value of Safety to better explain what we consider a risk of harm and why we remove content that could contribute to this risk.
  - Explained the relationship between the "policy rationales" and the standards themselves. This additional explanation aims to clarify the structure and intent of the Community Standards.
  - Clarified our Dangerous Individuals and Organizations policy to indicate that we allow discussion about the human rights of designated individuals or members of designated dangerous entities.
  - Published information about our approach to satirical content and added language explaining that satirical content will be allowed in seven sections of the Community Standards.
- **Undertook new policy development processes:**
  - Prepared a Crisis Policy Protocol in response to the board's May 2021 recommendation that we establish a policy to govern our response to crises. After research and consultation with internal and external stakeholders throughout the second half of the year, the policy was formally presented and adopted at Meta's Policy Forum in January 2022. The minutes of the Policy Forum session will soon be available in our Transparency Center.
  - Began preparations for a new policy development process to explore potentially adding religious and traditional allowances within our restricted goods and services policies.

In general, when we first review a recommendation from the board, we want to understand the potential effects of any action on our policies, systems, tools, and services. Our approach to research varies depending on the specific recommendation, but our overarching goal is the same – understanding impact. In our tests we might launch a version of a proposed product change, such as new messaging in a user notification to a subset of users. We then measure the effects on certain outcomes, such as fairness of the process according to user feedback, and the frequency of appeals. This fall, for example, we ran a test where we informed users whether automation or manual enforcement had resulted in their content being removed.

Similarly, we use the Policy Forum to discuss potential changes to our policies, including substantive updates that result from board recommendations. During the Policy Forum, subject matter experts from the Content Policy team propose adding new policies or amending existing

ones based on extensive research and consultation with external stakeholders. A variety of internal stakeholders provide input, including members from teams such as Safety and Cybersecurity Policy; Global Operations; Civil Rights and Human Rights; Legal; Communications; and Diversity, Equity, and Inclusion; as well as product managers and other public policy leads. We publish the minutes from these meetings here.

As we mentioned in our Q2 and Q3 2021 Quarterly Update, we've been exploring ways to share more information with the board around how we approach their recommendations. We plan to begin briefing the board with details of our work to implement its recommendations, including research results, on an ongoing basis. For example, we will brief the board about our product development cycle for launching more specific user messaging. We will discuss the design and results of specific experiments, and the metrics we use to decide whether to launch a product feature. And, in January, members of the board's staff attended the Policy Forum for the first time, to observe the deliberative process firsthand. Our goal is to increase transparency with the board about our processes and how we undertake the work for their recommendations.

Finally, we aim to provide as much detail as possible in our responses to the board's recommendations. We noted in our last quarterly update some of the challenges we faced when trying to provide thorough responses to the board's recommendations within 30 days. As the board noted in its most recent set of decisions, the Oversight Board bylaws were adjusted this month to extend our required response time for recommendations from 30 days to 60 days. This additional time will allow us to provide a more comprehensive response to the board's recommendations.

1. How to Read This Update

We designed this update in partnership with BSR (Business for Social Responsibility), based on best practices in human rights reporting principles, corporate disclosures, and goal-tracking reports. These include the Sustainability Accounting Standards Board (SASB) Conceptual Framework, International Integrated Reporting Council Framework, GRI Reporting Principles and UN Guiding Principles for Business and Human Rights, among others.

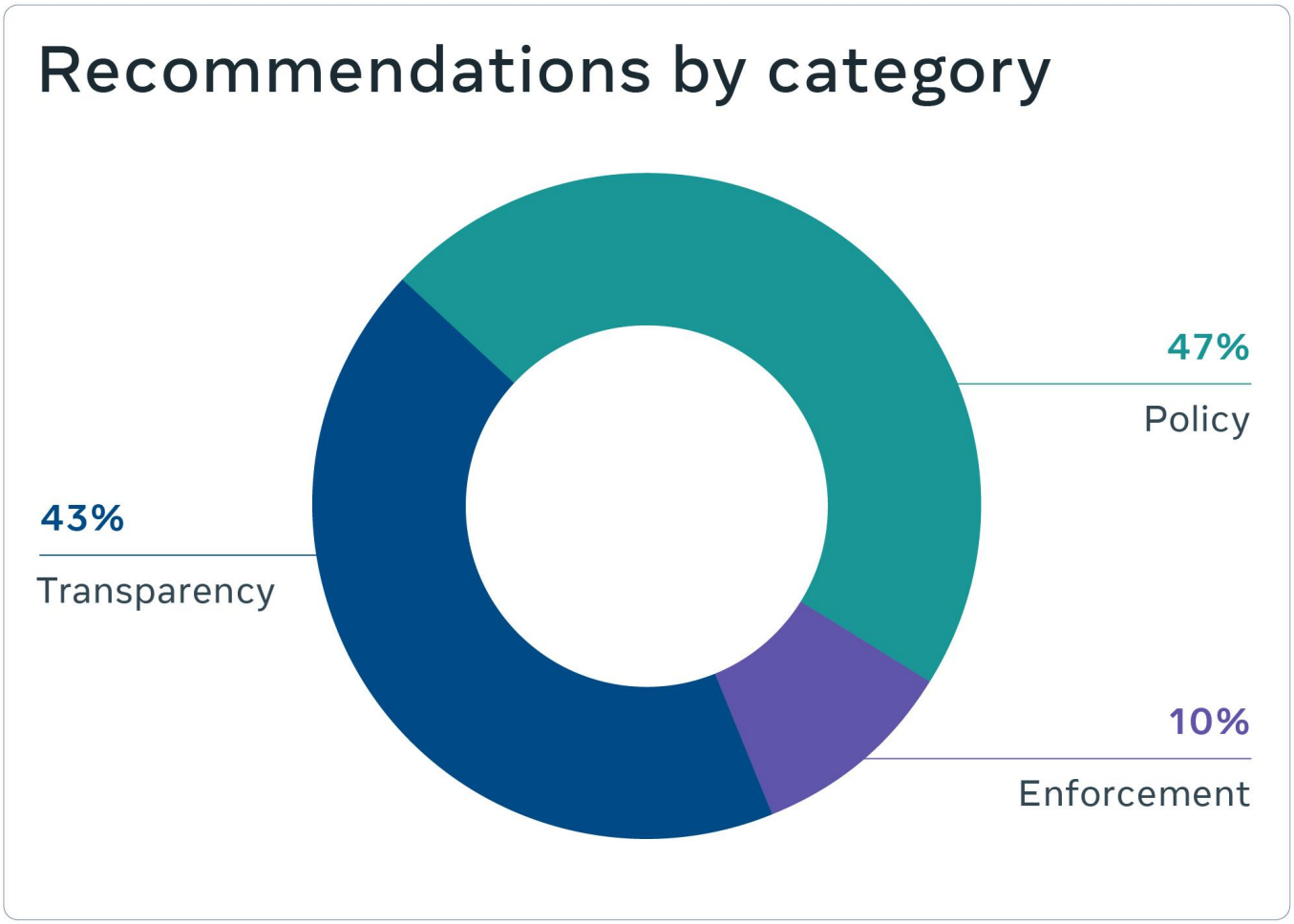
From January through December 2021, the board issued 87 non-binding recommendations. In our Q2 and Q3 2021 Quarterly Update, we addressed 69 of these recommendations and indicated we would provide more information on 40 in our next update. In this update we address those 40 recommendations as well as the 9 new recommendations the board included in the decisions it issued in Q4 of 2021. Out of the 9 new recommendations, we do not have updates for 2 because — as we explained in our 30-day response — we would either take no further action on the recommendation or it was work Meta already does.[3]

We are organizing our substantive updates on the 47 recommendations into three sections:

A. **Transparency (21 recommendations):** Helping people understand what our rules are, what violates them, and the consequences of violating them.
B. **Policy (22 recommendations):** Updating the Facebook Community Standards and Instagram Community Guidelines with new details and clearer explanations.
C. **Enforcement (4 recommendations):** Improving the quality and efficacy of our content moderation operations at scale.

Our goal is that these sections will help facilitate discussion around the progress made in these areas as well as improve the readability of this update. In the body of each section, we provide a general overview of our key areas of progress. For further detail on each recommendation, including the full text, please refer to the Appendix.

[3] This applies to recommendation #2 in the Post Discussing the Situation in Ethiopia case and recommendation #2 in the Depicting Indigenous Artwork and Discussing Residential Schools case. Future studies on user appeals, as described in Depicting Indigenous Artwork and Discussing Residential Schools Recommendation #2, will be tracked under Armenian People and the Armenian Genocide Recommendation #4



2. Update on Non-Binding Recommendations

A. Transparency

We are providing updates for our work on 21 board recommendations that address transparency and accountability.

We want to highlight our progress on:

1. **Completing the analysis of several proposals for creating more detailed notifications for users.**
    - **Automation vs. Human Review**: We ran a test where we informed users whether automation or human review resulted in their content being removed. Based on our findings, we are proceeding with plans to roll out this new user messaging in the first half of this year.

- **Violation Type**: We also measured the frequency and potential impact of instances where we change the violation type after reviewing a post a user has appealed to understand the potential benefit of updating user notifications. We will brief the board on the insights from these and other analyses of user messaging proposals in greater detail.

2. **Developing new messaging for users when their appeal to the Oversight Board resulted in Meta identifying and fixing our enforcement error**: We've made progress building this new notification feature, and will be working to implement the feature over the next few months.
3. **Testing more granular user notifications when enforcing our bullying & harassment policies**: Building on the progress we've made from the board's recommendations that we launch more specific messaging to people when they violate our Hate Speech Community Standards, we have also launched test versions of these notifications for people when their content goes against our Bullying & Harassment Community Standards, and expect to have results by the next Quarterly Update.

For a comprehensive list of all 21 recommendations in this category that we are providing updates on as part of this Quarterly Update, see Appendix A. Transparency.

B. Policy

We are providing updates for our work on 22 board recommendations that address our Facebook Community Standards and Instagram Community Guidelines.

We want to highlight our progress on:

1. **Undertaking new policy development**
   - **Crisis Protocol**: We prepared a detailed proposal for a Crisis Protocol, in accordance with a recommendation from the board, and presented it at our standing Policy Forum on January 25th, 2022. The proposal was adopted, and the new crisis protocol has been adopted. Minutes from the session will soon be found in our Transparency Center. Staff of the Oversight Board attended and observed these discussions.

- **Discussion of Non-Medical Drugs**: We have also begun to scope the policy development work in support of the board's recommendation that we add an allowance for the positive discussion of religious and traditional uses of non-medical drugs in our Restricted Goods & Services policies.

2. **Providing new guidance for content moderators reviewing potential violations of the Dangerous Individuals and Organizations policy**: Following a global consultation with Human Rights NGOs, activists, and academics, we outlined examples for both violating and non-violating examples of "support" for reviewers to consider when assessing content.
3. **Updating our Community Standards**
   - **Introduction**: We published a new explanation of the relationship between the "do nots" and the policy rationale in the introduction of the Community Standards.
   - **Safety Value**: We updated the "Safety" value in the Community Standards to reflect that online speech may pose risk to the physical security of persons.
   - **Dangerous Individuals and Organizations**: We described more clearly in our Dangerous Individuals and Organizations policy that we allow the discussion of human rights of designated individuals or members of designated dangerous entities.
   - **Satire Exception**: We updated our Community Standards to reflect that we allow satirical content in seven policy areas.

For a comprehensive list of all 22 recommendations in this category that we are providing updates on as part of this Quarterly Update, see Appendix B. Policy.

C. Enforcement

We are providing updates for our work on 4 board recommendations that address our enforcement systems.

We want to highlight our progress on:

1. **Sharing learnings from our early analyses for how Meta can give users more voice in appeals and improve appeals outcomes**: We have completed experiments giving users the ability to provide additional context with their appeals, and we found that receiving structured feedback may improve the accuracy of appeals review. We are planning to scope additional research this year.

2. **Identifying the challenges of considering more social and political context when reviewing potential bullying and harassment violations**: We completed a feasibility assessment that highlighted the operational complexity of defining “social” and “political” context at scale. We concluded that we could not consider additional social and political context at scale, but would continue to consider these contextual factors on escalation.

For a comprehensive list of all four recommendations in this category that we are providing updates on as part of this Quarterly Update, see Appendix C. Enforcement.

III. Appendix

How to Read This Appendix

The categorization of our responses to the board's recommendations include the following:

- **Implementing fully:** We agree with the recommendation and have or will implement it in full.
- **Implementing in part:** We agree with the overall aim of the recommendation and have or will implement work related to the board's guidance.
- **Assessing feasibility:** We are assessing the feasibility and impact of the recommendation.
- **No further action:** We will not implement the recommendation, for example, due to a lack of feasibility or disagreement about how to reach the desired outcome.
- **Work Meta* already does:** We have already addressed the recommendation through an action that we already do.

The current status for our responses to the board's recommendations include the following:

- **Complete**: We have completed full or partial implementation in line with our response to the board's recommendation, and will have no further updates on the recommendation.
- **In progress:** We are continuing to make progress on our response to the board's recommendation, and will have further updates on the recommendation.
- **No further updates:** We will not implement the recommendation or have already addressed the recommendation through an action that we already do, and will have no further updates on the recommendation.

*As we announced on October 28, 2021, Meta is the new name of the Facebook company. Therefore, all historical references to the Facebook company in this appendix refer to Meta.

Appendix A. Transparency

Transparency Recommendations	
Recommendation Text: *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.* *(Armenians in Azerbaijan Recommendation #1 (along with Breast Cancer Symptoms and Nudity Recommendation #3, Nazi Quote Recommendation #1, Depiction of Zwarte Piet Recommendation #2, South Africa Slur Recommendation #1, and Post Discussing a Substance with Psychoactive Properties Recommendation #2)[4])*	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
Feb. 2022 Update	We continue making progress on our ongoing work to provide clear, specific messaging to users. Since our Q2 and Q3 2021 Quarterly Update, we began to expand this work beyond hate speech policies, launching our first experiments with this type of specific messaging in instances of bullying & harassment policy violations, beginning with English, French, Hindi, Spanish, Arabic, Portuguese, and Indonesian. We anticipate completing this experimentation in the first quarter of this year. We've kept this commitment as "Implementing in Part" because we're building and testing these messages to see what potentially improves a person's experience, but we don't know yet if we'll launch these changes to all policy lines of the Community Standards. We expect to continue testing the expansion of these more specific notifications to other policy lines. Because this work is gradual and often does not result in immediate, observable product changes, we are providing the board an in-depth briefing on our ongoing implementation of these product recommendations.
Recommendation Text: *Inform users when automation is used to take enforcement action against their content, including accessible descriptions of what this means.* *(Breast Cancer Symptoms & Nudity Recommendation #5)*	
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	In Progress

[4] The board issued similar recommendations in the following cases: Breast Cancer Symptoms and Nudity #3, Nazi Quote #1, Depiction of Zwarte Piet #2, South Africa Slur #1, and Post Discussing a Substance with Psychoactive Properties #2. We are tracking the progress of our work in response to these recommendations as part of our response to recommendation #1 in the Armenians in Azerbaijan case.

Feb. 2022 Update	From September through November 2021, we ran an experiment for a small set of people in which we informed them whether automation or human review led to their content being taken down. We analyzed how this affected people's experiences (such as whether the process was fair) and behaviors (such as whether the number of appeals or subsequent violations decreased). We've decided to proceed with launching this new user messaging. We will start in certain locations to further understand the effects of this message for different policies and user groups, and iterate on our design to ensure user comprehension. We will continue to monitor how this affects people's experiences to ensure that the goals of this recommendation are being achieved, and brief the board on our findings. We've kept this commitment as "Implementing in Part" because of the gradual roll-out and we will consider changing this to "Implementing Fully" if the roll out is widespread and achieving the impact we are hoping for. We anticipate starting to roll out this new messaging in the first half of this year.
Recommendation Text: *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.* *(Breast Cancer Symptoms & Nudity Recommendation #6)*	
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	We continue to explore ways of expanding our transparency reporting to present data on our automated enforcement. However, because of the complexity of the systems and the scope of the expansion, it will take time to determine the most appropriate way to measure this metric. Aggregating enforcement actions under the category of "automated" or "manual" review is not always straightforward, because many decisions are made with a combination of both manual and automated input. For example, a piece of content might be removed by an automated classifier and that decision may be confirmed by a human reviewer on appeal. We are determining appropriate ways of accounting for all scenarios in potential metrics, and we will provide progress updates in subsequent Quarterly Updates.
Recommendation Text: *Facebook should more clearly explain its newsworthiness allowance.* *(Former President Trump's Suspension Recommendation #11)*	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress

Feb. 2022 Update	We are exploring how to build the systems and processes to measure and publish the number of times we granted newsworthiness allowances. In addition, in line with our commitment in Post Depicting Protests in Colombia While Using a Slur Recommendation #2, we anticipate sharing examples of our newsworthiness allowance in the Transparency Center, and we will track further progress on that work under this recommendation.
Recommendation Text: *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.* *(Former President Trump's Suspension Recommendation #18)*	
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	In Progress
Feb. 2022 Update	We are actively working on two initiatives resulting from this recommendation. The first initiative, and our current priority, is to measure our enforcement actions on accounts, Pages, and Groups. Measuring enforcement actions on these kinds of entities is more complicated than for single pieces of content such as photos, videos, or posts. For example, multiple types of content violations might contribute to the removal of a single account. The second initiative is to measure enforcement data by location. Attributing a given content enforcement action to a specific location is complicated. For example, if users in multiple countries post the same piece of violating content, or users in multiple countries report the same piece of content, a challenge arises when trying to determine which location we should attribute the violation or removal to. In addition, some users - particularly those who violate certain Community Standards, such as spam - conceal their IP address to hide their location. Taken together, these definitional questions and the uncertainty in the data present significant location attribution challenges. Although we are working on both how to measure enforcement actions on accounts, Pages, and Groups, and how to represent enforcement data by location, we cannot commit to publishing these metrics in 2022. Our current priorities are to define the metrics and create processes for measuring them before we report them publicly. We will provide progress updates on this work in future Quarterly Updates.
Recommendation Text: *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.* *(Punjabi Concern Over the RSS in India Recommendation #3)*	
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	In Progress

Feb. 2022 Update	Since our last update, we've started work toward the goal of sharing information about our enforcement error rates across the different sections of the Community Standards. We anticipate beginning work to build accuracy metrics this year, but given existing priorities on our data reporting roadmaps (such as those outlined in our response to Former President Trump's Suspension Recommendation #18), we do not expect to complete this work in 2022. In addition, as we explain in the update to Former President Trump's Suspension Recommendation #18, we are actively working on the challenges of measuring enforcement data by location. We will provide additional updates in future Quarterly Updates.

Recommendation Text: *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.*

(January 2021 Protests in Russia Recommendation #6)

Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	In our initial response to this recommendation, we described existing product features that provide users the opportunity to modify or remove their violating post before we enforce our policies. These "self-remediation" features are now live on Facebook for multiple policy violations, including bullying and harassment, hate speech, adult nudity and sexual activity, and graphic violence, and we expect to expand next to violence and incitement. The notifications for posts in bullying and harassment are currently available in Arabic, Spanish, English, Indonesian, Portuguese, Hindi, Burmese, German, Urdu, Bengali, French, Thai, Russian, Tagalog, and Italian. We inform users of the violation at the policy level (such as "bullying and harassment"), rather than at the more specific rule level (such as "negative character claim") because our automated systems are the most accurate at the policy level, and we do not want to risk sending users incorrect or confusing messaging. In addition, we aren't able to provide this level of immediate, specific notice from our manual review systems because of the volume of content we review. Because we believe our current product features and ongoing investments address the core of the board's recommendation, there will be no further updates on this recommendation.

Recommendation Text: *Facebook should make technical arrangements to ensure that notice to users refers to the Community Standard enforced by the company. If a user's content violates a different Community Standard to the one the user was originally told about, they should have another opportunity to appeal.*

(Armenian People and the Armenian Genocide Recommendation #1)

Previous Category	Assessing Feasibility
Updated Category	No Further Action

Current Status	No Further Updates
Feb. 2022 Update	We analyzed our appeals data to understand how often a violation type changes as a result of reviewing an enforcement decision on appeal, and what that changed enforcement might mean for user experiences. Based on our preliminary analysis, we found only a small percentage of appeals where the violation type changed on appeal. This typically occurred between similar policy areas, for example, a violation changing from "Violence and Incitement" to "Dangerous Individuals and Organizations." In order to accommodate this particular instance of user messaging, we would need to be able to attribute multiple policy violations to a single piece of content, which would require fundamental, system-wide changes to the design of our content review systems. Because our analysis found that this occurred only in a small percentage of appeals, we are prioritizing other product solutions to increase the quality of user experiences. There will be no further updates on this recommendation.
Recommendation Text: *Ensure that users are notified when their content is removed. The notification should note whether the removal is due to a government request or due to a violation of the Community Standards or due to a government claiming a national law is violated (and the jurisdictional reach of any removal).* *(Support of Abdullah Öcalan, Founder of the PKK Recommendation #9)*	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	As we describe in Support of Abdullah Ocalan Recommendation #11, we have begun building new infrastructure that will enable this measurement and reporting. Building new messages for users about these government requests is dependent on the launch of these updated internal systems. Although we are not planning to launch this new user messaging in 2022, we expect to begin assessing options and scoping work for it later this year. We will provide more information in a future Quarterly Update.
Recommendation Text: *Include information on the number of requests Facebook receives for content removals from governments that are based on Community Standards violations (as opposed to violations of national law), and the outcome of those requests.* *(Support of Abdullah Öcalan, Founder of the PKK Recommendation #11 (along with Al Jazeera Post on Tensions Between Israel and Palestine #4[5])*	
Previous Category	Implementing Fully
Updated Category	Implementing Fully

[5] The board issued a similar recommendation in recommendation #4 in the Al Jazeera Post on Tensions Between Israel and Palestine case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #11 in the Support of Abdullah Öcalan case.

Current Status	In Progress
Feb. 2022 Update	We have completed scoping the work to develop the data infrastructure for including information about the requests we receive through our standardized global process for government requests for content removals in our transparency reporting. Our priority is in building tools and processes that enable us to report this information across the diversity of request formats we receive, with the goal of producing five country-level metrics, including the number of: 1. unique requests we receive; 2. pieces of content covered by these requests; 3. such pieces of content removed under the Community Standards; 4. such pieces of content locally restricted based on local law; and 5. such pieces of content where no action is taken. Because this is a large, complex project, we do not anticipate publishing the new reports in 2022. We expect to complete the initial changes to internal infrastructure this year, and will provide an update on timeline for public reporting in a future Quarterly Update.

Recommendation Text: *Include more comprehensive information on error rates for enforcing rules on "praise" and "support" of dangerous individuals and organizations, broken down by region and language.*

(Support of Abdullah Öcalan, Founder of the PKK Recommendation #12)

Previous Category	Assessing Feasibility
Updated Category	No Further Action
Current Status	No Further Updates
Feb. 2022 Update	Since our previous update, we have determined that we will not include enforcement data reports at the level of granularity this recommendation outlines. We are instead prioritizing the work that will enable broader, report-level changes, such as publishing enforcement data on complex objects and by location (see update to Former President Trump's Suspension Recommendation #18, above). There will be no further updates on this recommendation.

Recommendation Text: *Engage an independent entity not associated with either side of the Israeli-Palestinian conflict to conduct a thorough examination to determine whether Facebook's content moderation in Arabic and Hebrew, including its use of automation, have been applied without bias. The report and its conclusions should be made public.*

(Al Jazeera Post on Tensions Between Israel and Palestine Recommendation #3)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress

Feb. 2022 Update	Independent human rights due diligence began mid-2021, and is ongoing. Due to the project's complexity, we regret that we will not have insights and actions to share until the next Quarterly Update, rather than in the first quarter as we previously had hoped.

Recommendation Text: *Notify all users who reported content assessed as violating but left on the platform for public interest reasons that the newsworthiness allowance was applied to the post. The notice should link to the Transparency Center explanation of the newsworthiness allowance.*

(Post Depicting Protests in Colombia While Using a Slur Recommendation #4)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	We are continuing to explore how to build product features that we can use to inform users when a piece of content that violates the Community Standards is left up under our newsworthiness allowance. We plan to provide additional information in a future Quarterly Update.

Recommendation Text: *Provide users with timely and accurate notice of action being taken on the content their appeal relates to. Where applicable, including in enforcement error cases like this one, the notice to the user should acknowledge that the action was a result of the Oversight Board's review process. Meta should share the user messaging sent when board actions impact content decisions appealed by users, to demonstrate it has complied with this recommendation.*

(Depicting Indigenous Artwork and Discussing Residential Schools Recommendation #1)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	As described in our January 7, 2022 response, we will be updating the messaging that users receive when we notify them of a change to the status of their content due to an enforcement error detected as a result of the user's appeal to the board. The messaging will read: "As a result of your Oversight Board appeal, we reviewed your [content] again and found that we removed it by mistake. We have now restored your [content], as it did not violate our Community Standards. We're sorry we got this wrong. We're looking into what went wrong and will continue to improve how we detect and remove content." We will be working to implement this change over the course of the next few months, and will share an update in a future Quarterly Update when it is complete.

Recommendation Text: *Commission an independent human rights due diligence assessment on how Facebook and Instagram have been used to spread hate speech and unverified rumors that heighten the risk of violence in Ethiopia. The assessment should review the success of measures Meta took to prevent the misuse of its products and services in Ethiopia. The assessment should also review the success of measures Meta took to allow for corroborated and public interest reporting on human rights atrocities in Ethiopia. The assessment should review Meta's language capabilities in Ethiopia and if they are adequate to protect the rights of its users. The assessment should cover a period from June 1, 2020, to the present. The company should complete the assessment within six months from the moment it responds to these recommendations. The assessment should be published in full.*

(Post Discussing the Situation in Ethiopia Recommendation #3)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	In our January 13, 2022 response, we committed to continuing existing human rights due diligence and dynamic risk management processes related to Ethiopia. Not all elements of the board's recommendation may be feasible in terms of timing, data science, or approach, but we will work to update and share insights and actions from our due diligence that aligns with the board's goals, our Human Rights Policy, and the UN Guiding Principles on Business and Human Rights. We anticipate providing an update in the first half of this year.

Appendix B. Policy

Policy Clarity & Accessibility Recommendations	
Recommendation Text: *Revise the Instagram Community Guidelines to specify that female nipples can be shown to raise breast cancer awareness and clarify that where there are inconsistencies between the [Instagram] Community Guidelines and the [Facebook] Community Standards, the latter take precedence.* *(Breast Cancer Symptoms & Nudity Recommendation #2 (along with Support of Abdullah Öcalan Recommendation #10, and Post Discussing a Substance with Psychoactive Properties Recommendation #1)[6])*	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	In our 30-day response to Post Discussing a Substance with Psychoactive Properties Recommendation #1, we committed to publishing updates to the Instagram Community Guidelines so they match the Facebook Community Standards in all of the shared policy areas; and to make the differences clear in the small number of instances where the policies differ. Our work is currently underway, and we anticipate its completion in the coming months.
Recommendation Text: *Facebook should be clear in its Corporate Human Rights Policy how it collects, preserves, and shares information related to investigations and potential prosecutions, including how researchers can access that information.* *(Former President Trump's Suspension Recommendation #15)*	
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	In Progress
Feb. 2022 Update	In our previous update, we explained that we would review our Corporate Human Rights Policy for opportunities to improve our protocols for information collection, preservation and sharing. We already explicitly reference Meta's law enforcement guidelines and data policy, which are the two existing policies most germane to criminal investigations and prosecutions. We will engage the board to provide more in-depth briefings on our current and evolving frameworks relating to their recommendation. In addition, we will update corporate human rights policy language on a periodic basis and include language on any new relevant policy areas at that time.

[6] The board issued a similar recommendation in recommendation #10 in the Support of Abdullah Öcalan case and recommendation #1 in the Post Discussing a Substance with Psychoactive Properties case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #2 in the Breast Cancer Symptoms & Nudity case.

Recommendation Text: *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.* (*Former President Trump's Suspension Recommendation #19*)	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	The Crisis Policy Protocol was formally presented and adopted at Meta's Policy Forum in January 2022 with Oversight Board staff in attendance. Protocol development included original research, consultations with over 50 global external experts in national security, conflict prevention, hate speech, humanitarian response, and human rights, and over six internal working groups representing a range of interdisciplinary expertise in crisis response from across product, operational and policy teams. Per our existing transparency practices, this quarter we will publish the Policy Forum deck providing an overview of the Crisis Policy Protocol in the Transparency Center.
Recommendation Text: *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.* (*Punjabi Concern Over the RSS in India Recommendation #1*)	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	Because of the board's recommendation, we prioritized translating the Community Standards into five additional Indic languages. We published the Community Standards in Gujarati, Marathi, Punjabi, Tamil and Telugu. This brings our Community Standards coverage to 59 languages and regions, a 20% increase just within this half. There will be no further updates on this recommendation.
Recommendation Text: *Explain the relationship between its Bullying and Harassment policy rationale and the "Do nots" as well as the other rules restricting content that follow it.* (*January 2021 Protests in Russia Recommendation #1*)	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete

Feb. 2022 Update	To clarify and explain the relationship between the policy rationale and the sections that follow it, we have introduced the following language into the introduction of the Community Standards: "Each section of our Community Standards starts with a "Policy Rationale" that sets out the aims of the policy followed by specific policy lines that outline: • Content that is not allowed; and • Content that requires additional information or context to enforce on, content that is allowed with a warning screen or content that is allowed but can only be viewed by adults aged 18 and older." There will be no further updates on this recommendation.

Recommendation Text: *Differentiate between bullying and harassment and provide definitions that distinguish the two acts. The Community Standard should also clearly explain to users how bullying and harassment differ from speech that only causes offense and may be protected under international human rights law.*

(January 2021 Protests in Russia Recommendation #2)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	We are continuing to conduct research that evaluates how bullying and harassment may be similar or different in order to inform our Bullying and Harassment Community Standards. Development and changes to the Community Standards often include consideration from a range of inputs, including research, external expertise, and internal feedback. We anticipate having more to share in the next Quarterly Update.

Recommendation Text: *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.*

(January 2021 Protests in Russia Recommendation #3)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	In our Q2 and Q3 2021 Quarterly Update we shared that, in response to this recommendation, we updated the Bullying and Harassment section of our Community Standards to clarify our policy as it relates to private individuals, involuntary public figures and public figures who are minors. We shared these changes as part of our August 2021 Community Standards updates, which included the addition of tiers and language clarifications. We are continuing to explore ways to further share definitions and examples of the Bullying and Harassment section of the Community Standards in support of this recommendation, and will share additional updates in a future Quarterly Update.

Recommendation Text: *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.* (*January 2021 Protests in Russia Recommendation #4*)	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
Feb. 2022 Update	As we explained in our 30-day response, we reviewed our internal policy guidance to ensure it adequately provides examples of violating and non-violating content. We are continuing to explore ways to share more details about our enforcement approach to Bullying and Harassment, and will share further updates in a future Quarterly Update. We will track future progress on this work as part of our updates to January 2021 Protests in Russia Recommendation #3, above.
Recommendation Text: *To make its policies and their enforcement clearer for users, Facebook should: Include the exception, which is currently not communicated to users, in the public language of the Hate Speech Community Standard.* (*Armenian People and the Armenian Genocide Recommendation #2*)	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	Complete
Feb. 2022 Update	In November we updated the Community Standards to reflect that satirical content will be allowed when we have the necessary context in the following sections: • Dangerous Individuals and Organizations • Restricted Goods and Services (formerly Regulated Goods) • Fraud and Deception • Hate Speech • Privacy Violations (where information is fake) • Adult Nudity and Adult Sexual Activity • Adult Sexual Solicitation and Sexually Explicit Language There will be no further updates on this recommendation.
Recommendation Text: *To improve the accuracy of the enforcement of its content policies for the benefit of users, Facebook should: Make sure that it has adequate procedures in place to assess satirical content and relevant context properly. This includes providing content moderators with: (i) access to Facebook's local operation teams to gather relevant cultural and background information; and (ii) sufficient time to consult with Facebook's local operation teams and to make the assessment. Facebook should ensure that its policies for content moderators incentivize further investigation or escalation where a content moderator is not sure if a meme is satirical or not.* (*Armenian People and the Armenian Genocide Recommendation #3*)	

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	In preparing the update to the Community Standards to reflect the satire context specific policy, we also developed procedures and guidelines to support escalations teams in evaluating content for satire. As in other areas of policy, these guidelines were developed with careful consideration of safety, voice, and equity. There will be no further updates on this recommendation.
Recommendation Text: *Publish the results of the ongoing review process to determine if any other polices were lost, including descriptions of all lost policies, the period the policies were lost for, and steps taken to restore them.* (*Support of Abdullah Öcalan, Founder of the PKK Recommendation #3*)	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	Last half, we completed a review of internal guidance related to this recommendation to ensure that this guidance was up to date and consistent. Our review confirmed that all of our established policies are being correctly implemented by our reviewers. However, we did identify, and fix, two instances of outdated guidance in training materials. We assessed little to no impact to content moderation decisions because the relevant topics are addressed in more detail in more recent reviewer guidance. We are taking steps to ensure that the most up to date guidance is in place moving forward. There will be no further updates on this recommendation.
Recommendation Text: *Reflect in the Dangerous Individuals and Organizations "policy rationale" that respect for human rights and freedom of expression can advance the value of "Safety," and that it is important for the platform to provide a space for these discussions.* (*Support of Abdullah Öcalan, Founder of the PKK Recommendation #4*)	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	Complete

Feb. 2022 Update	In December, we updated our Dangerous Individuals and Organizations policy language to clarify that we allow discussion about the human rights of designated individuals or members of designated dangerous entities when that content does not include other praise, substantive support, or representation of designated entities or other policy violations. In that update, we also included a link for users to review our Corporate Human Rights Policy to learn more about our commitments to internationally recognized human rights. There will be no further updates on this recommendation.
Recommendation Text: *Add to the Dangerous Individuals and Organizations policy a clear explanation of what "support" excludes. Users should be free to discuss alleged violations and abuses of the human rights of members of designated organizations. Calls for accountability for human rights violations and abuses should also be protected.* *(Support of Abdullah Öcalan, Founder of the PKK Recommendation #5)*	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	Complete
Feb. 2022 Update	In addition to including an explanation about allowing discussion of human rights in the Dangerous Individuals and Organizations section of the Community Standards, we have also outlined examples of both violating and non-violating explanations of "support" for reviewers to consider when assessing content. We conducted a global consultation with Human Rights NGOs, activists, and academics to guide the drafting process of this change. As such, guidance was crafted with consideration of UN Human Rights conventions. There will be no further updates on this recommendation.
Recommendation Text: *Explain in the Community Standards how users can make the intent behind their posts clear to Facebook. This would be assisted by implementing the Board's existing recommendation to publicly disclose the company's list of designated individuals and organizations (see: case 2020-005-FB-UA). Facebook should also provide illustrative examples to demonstrate the line between permitted and prohibited content, including in relation to the application of the rule clarifying what "support" excludes.* *(Support of Abdullah Öcalan, Founder of the PKK Recommendation #6)*	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	In response to Support of Abdullah Öcalan, Founder of the PKK Recommendation #5, illustrative examples have now been provided to reviewers to clarify the line of "support". We explained in our previous Quarterly Update that, because of the potential safety risks to our teams and tactical challenges to our ability to stay ahead of adversarial shifts, we determined not to publish any additional detail about the designations in this policy area. There will be no further updates on this recommendation.

Recommendation Text: *Add criteria and illustrative examples to its Dangerous Individuals and Organizations policy to increase understanding of the exceptions for neutral discussion, condemnation and news reporting.* *(Al Jazeera Post on Tensions Between Israel and Palestine Recommendation #1)*	
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	As a result of this recommendation, subject matter experts are exploring ways to add more clarity to our Dangerous Individuals and Organizations policy and are committed to providing an explanation regarding exceptions for neutral discussion, condemnation, and news reporting in this policy section. There is also ongoing work to examine the potential of including key examples for these exceptions in the Community Standards. We will provide further updates on the progress of implementing these examples in a future Quarterly Update.
Recommendation Text: *Ensure swift translation of updates to the Community Standards into all available languages.* *(Al Jazeera Post on Tensions Between Israel and Palestine Recommendation #2)*	
Previous Category	Assessing Feasibility
Updated Category	No Further Action
Current Status	No Further Updates
Feb. 2022 Update	In our last Quarterly Update, we committed to exploring whether we could shorten the turnaround time for translating updates to our Community Standards in all available languages. Currently, translations in a given language take roughly a week, but it can take up to four to six weeks to complete translations to all available languages, depending on the length and complexity of the update. Our review determined that potential changes to the current process would not result in faster turnaround time for translations. There will be no further updates on this recommendation.
Recommendation Text: *Publish illustrative examples from the list of slurs it has designated as violating under its Hate Speech Community Standard. These examples should be included in the Community Standard and include edge cases involving words which may be harmful in some contexts but not others, describing when their use would be violating. Facebook should clarify to users that these examples do not constitute a complete list.* *(Post Depicting Protests in Colombia While Using a Slur Recommendation #1)*	
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part

Current Status	In Progress
Feb. 2022 Update	We are working to update the explanation of our slurs policy in the Community Standards. Because these terms may create an environment of intimidation and exclusion, we have determined not to publish a comprehensive list of violating slurs. However, internal teams comprising subject matter expertise and regional knowledge are identifying examples of slurs with the aim of illustrating certain words that may be harmful in some cases but not others. We anticipate publishing these examples and will make clear that it does not constitute a complete list.

Recommendation Text: *Link the short explanation of the newsworthiness allowance provided in the introduction to the Community Standards to the more detailed Facebook Transparency Center explanation of how this policy applies. The company should supplement this explanation with illustrative examples from a variety of contexts, including reporting on large scale protests.*

(Post Depicting Protests in Colombia While Using a Slur Recommendation #2)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
Feb. 2022 Update	As stated in our Q2 and Q3 2021 Quarterly Update, we committed to including a link to our explanation of our approach to newsworthy content in the introduction of the Community Standards. We are working to identify potential examples to share publicly to provide further explanation of the application of the newsworthiness allowance, and will provide updates on the progress of this initiative in future Quarterly Updates. We will track future progress on this recommendation as part of our work to explore how to build the systems and processes to measure and publish the number of times we granted newsworthiness allowances in Former President Trump's Suspension Recommendation #11.

Recommendation Text: *The board recommends that Meta modify the Instagram Community Guidelines and Facebook Regulated Goods Community Standard to allow positive discussion of traditional and religious uses of non-medical drugs where there is historic evidence of such use. The board also recommends that Meta make public all allowances, including existing allowances.*

(Post Discussing a Substance with Psychoactive Properties Recommendation #3)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress

Feb. 2022 Update	In our January 7, 2022 response, we shared that we are assessing the feasibility of this recommendation through our standard policy development process, including as a discussion in the Policy Forum. The policy team has begun initial conversations and scoping to start this policy development in the upcoming months. As with past Policy Forum development, the team intends to consult with external stakeholders, conduct research, and gather input from a range of teams at Meta to inform recommendations for potential changes to the Restricted Goods and Services policy.
Recommendation Text: *Rewrite its value of "Safety" to reflect that online speech may pose risk to the physical security of persons and the right to life, in addition to the risks of intimidation, exclusion and silencing.* *(Post Discussing the Situation in Ethiopia Recommendation #1)*	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
Feb. 2022 Update	In January 2022, we updated our value of Safety in the Community Standards to reflect the board's recommendation. We will continue to adjust our policies and values as needed to mitigate any risk of contributing to offline harm, and continue to welcome guidance from the board to help us in these efforts. There will be no further updates on this recommendation.

Appendix C. Enforcement

Enforcement Recommendations	
Recommendation Text: *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.* (*January 2021 Protests in Russia Recommendation #5*)	
Previous Category	Assessing Feasibility
Updated Category	No Further Action
Current Status	No Further Updates
Feb. 2022 Update	In December 2021, we concluded an assessment of whether to require a consideration of the social and political context when reviewing potential bullying & harassment violations at scale. We determined that it would not be possible to consistently and comprehensively define signals of "social" and "political" context at scale. Attempting to broaden either our human review guidelines or our automated detection systems to accommodate this kind of context would introduce too much subjectivity, and result in inconsistent enforcement. Although the qualitative, definitional challenges concerning social and political context are a limiting factor at the level of scaled review, we do have protocols for assessing social and political context in certain circumstances. For example, our internal escalations teams routinely make content decisions that incorporate social and political context. We have numerous protocols to support this higher-touch, context-based content review and ensure it is objective. For example, Meta regularly establishes an Integrity Product Operations Center ("IPOC"), which is a working group composed of subject matter experts from our product, policy, and operations teams. This structure allows these experts to more quickly surface, triage, investigate, and mitigate risks on the platform. We have prepared a report of this feasibility assessment, and will share the findings of this assessment with the board in greater detail. There will be no further updates on this recommendation.
Recommendation Text: *Facebook should let users indicate in their appeal that their content falls into one of the exceptions to the Hate Speech policy.* (*Armenian People and the Armenian Genocide Recommendation #4*)[7]	
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility

[7] In our Q2 and Q3 2021 Quarterly Update, we categorized recommendation #4 in the Armenian People and the Armenian Genocide case as a Transparency recommendation. However, we have re-categorized this recommendation under Enforcement.

Current Status	In Progress
Feb. 2022 Update	We are actively exploring ways of giving users more voice in the appeals experience. In our 30-day response to recommendation #2 in the Depicting Indigenous Artwork and Discussing Residential Schools case, we described two previous experiments in which we gave users the ability to provide additional context with their appeal. In the first experiment, we provided users a free-form text box, and in the second we also offered a more structured drop-down menu. We explained that the results of this experiment suggested that allowing people to provide additional context on appeal could have an impact on enforcement outcomes. In the first half of 2022, we expect to begin designing a third experiment that builds on what we've learned from this prior work. For instance, one design option is to offer users a more detailed dropdown menu depending on the violation type, such as a specific dropdown menu with options for Hate Speech exceptions. We will discuss our progress in a future Quarterly Update.

Recommendation Text: *To improve the accuracy of Facebook's review in the appeals stage, the company should ensure appeals based on policy exceptions are prioritized for human review.*

(Armenian People and the Armenian Genocide Recommendation #5)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
Feb. 2022 Update	We currently review the vast majority of appeals manually. We are planning to assess potential changes to our appeals process that would prioritize certain content more urgently. These changes would be large, requiring a system-wide update, so we need to scope the requirements and assess the benefits before potentially implementing this recommendation. An additional factor in this type of system-design includes user-provided input in our review of appeals, such as the work we are assessing as a result of Armenian People and the Armenian Genocide Recommendation #4. We are in the early stages of defining our priorities in this space, with user research initiatives and data analysis underway. We will discuss our progress in a future Quarterly Update.

Recommendation Text: *Conduct accuracy assessments focused on Hate Speech policy allowances that cover artistic expression and expression about human rights violations (e.g., condemnation, awareness raising, self-referential use, empowering use). This includes how the location of a reviewer impacts the ability of moderators to accurately assess hate speech and counter speech from the same or different regions. Meta should share the results of this assessment with the board, including how these results will inform improvements to enforcement operations and policy development and whether it plans to run regular reviewer accuracy assessments on these allowances, and summarize the results in its Quarterly Updates.*

(Depicting Indigenous Artwork and Discussing Residential Schools Recommendation #3)

Previous Category	Assessing Feasibility
Updated Category	Implementing in Part
Current Status	In Progress
Feb. 2022 Update	Since our 30-day response, we have explored the question of whether the location of reviewers impacts their ability to accurately assess pieces of potential hate speech content from the same or different regions. We ran an experiment to test this hypothesis and plan to brief the board on our findings. In addition, as we described in our 30-day response to this recommendation, our automated and manual review systems do not currently identify reasons for why a piece of content is left up as non-violating. In order to conduct robust accuracy assessments of policy allowances, we need to indicate the reason for the allowance on a set of pieces of content. We are continuing to assess whether to run an analysis on the basis of limited data that is currently available, and whether we can create a dataset large enough for this analysis given the challenges associated with identifying these edge cases. We are also exploring more system-level options for better understanding the accuracy rates of how we apply policy allowances. We will continue to assess the feasibility of this work, and provide more information in a future Quarterly Update.

IV. Index

Case	Recommendation	Updated Category	Status	Section	Page
Armenians in Azerbaijan	2020-003-FB-UA-1	Implementing in part	In progress	Transparency	16
Breast cancer symptoms and nudity	2020-004-IG-UA-2	Implementing fully	In progress	Policy	24
	2020-004-IG-UA-3	Implementing in part	In progress	Transparency Footnote 4	16
	2020-004-IG-UA-5	Implementing in part	In progress	Transparency	16
	2020-004-IG-UA-6	Assessing feasibility	In progress	Transparency	17
Nazi quote	2020-005-FB-UA-1	Implementing in part	In progress	Transparency Footnote 4	16
Former President Trump	2021-001-FB-FBR-11	Implementing fully	In progress	Transparency	17
	2021-001-FB-FBR-15	Implementing in part	In progress	Policy	24
	2021-001-FB-FBR-18	Implementing in part	In progress	Transparency	18
	2021-001-FB-FBR-19	Implementing fully	In progress	Policy	25
Depiction of Zwarte Piet	2021-002-FB-UA-2	Implementing in part	In progress	Transparency Footnote 4	16
Punjabi Concerns Over the RSS in India	2021-003-FB-UA-1	Implementing in part	Complete	Policy	25
	2021-003-FB-UA-3	Implementing in part	In progress	Transparency	18
January 2021 Protests in Russia	2021-004-FB-UA-1	Implementing in part	Complete	Policy	25
	2021-004-FB-UA-2	Assessing feasibility	In progress	Policy	26
	2021-004-FB-UA-3	Implementing fully	In progress	Policy	26
	2021-004-FB-UA-4	Implementing in part	In progress	Policy	27
	2021-004-FB-UA-5	No further action	No further updates	Enforcement	33
	2021-004-FB-UA-6	Implementing in part	Complete	Transparency	19
Armenian People and the Armenian Genocide	2021-005-FB-UA-1	No further action	No further updates	Transparency	19
	2021-005-FB-UA-2	Implementing fully	Complete	Policy	27
	2021-005-FB-UA-3	Implementing in part	Complete	Policy	27
	2021-005-FB-UA-4	Assessing feasibility	In progress	Enforcement	33
	2021-005-FB-UA-5	Assessing feasibility	In progress	Enforcement	34

Support of Abdullah Ocalan	2021-006-IG-UA-3	Implementing in part	Complete	Policy	28
	2021-006-IG-UA-4	Implementing fully	Complete	Policy	28
	2021-006-IG-UA-5	Implementing fully	Complete	Policy	29
	2021-006-IG-UA-6	Implementing in part	Complete	Policy	29
	2021-006-IG-UA-9	Implementing fully	In progress	Transparency	20
	2021-006-IG-UA-10	Implementing fully	In progress	Policy Footnote 6	24
	2021-006-IG-UA-11	Implementing fully	In progress	Transparency	20
	2021-006-IG-UA-12	No further action	No further updates	Transparency	21
Al Jazeera Post on Tensions Between Israel and Palestine	2021-009-FB-UA-1	Assessing feasibility	In progress	Policy	30
	2021-009-FB-UA-2	No further action	No further updates	Policy	30
	2021-009-FB-UA-3	Implementing fully	In progress	Transparency	21
	2021-009-FB-UA-4	Implementing fully	In progress	Transparency Footnote 5	20
Post Depicting Protests in Colombia While Using a Slur	2021-010-FB-UA-1	Implementing in part	In progress	Policy	30
	2021-010-FB-UA-2	Implementing fully	In progress	Policy	31
	2021-010-FB-UA-4	Assessing feasibility	In progress	Transparency	22
South Africa Slur	2021-011-FB-UA-1	Implementing in part	In progress	Transparency Footnote 4	16
Depicting Indigenous Artwork and Discussing Residential Schools	2021-012-FB-UA-1	Implementing fully	In progress	Transparency	22
	2021-012-FB-UA-2	Work Meta already does	No further updates	Footnote 3	10
	2021-012-FB-UA-3	Implementing in part	In progress	Enforcement	34
Post Discussing a Substance with Psychoactive Properties	2021-013-IG-UA-1	Implementing fully	In progress	Policy Footnote 6	24
	2021-013-IG-UA-2	Implementing in part	In progress	Transparency Footnote 4	16
	2021-013-IG-UA-3	Assessing feasibility	In progress	Policy	31
Post Discussing the Situation in Ethiopia	2021-014-FB-UA-1	Implementing in part	Complete	Policy	32
	2021-014-FB-UA-2	No further action	No further updates	Footnote 3	10
	2021-014-FB-UA-3	Assessing feasibility	In progress	Transparency	23