Meta Q1 2022 Quarterly Update on the Oversight Board



TABLE OF CONTENTS

Intro	ntroduction	
I.	Meta's Referrals	4
II.	Progress on Non-Binding Recommendations	7
	How to Read This Update	9
	Non-Binding Recommendation Implementation Highlights	11
	Transparency	11
	Policy	12
	Enforcement	14
III.	Appendix	15
	How to Read This Appendix	15
	Appendix A. Transparency	16
	Appendix B. Policy	23
	Appendix C. Enforcement	31
IV.	Index	36

Introduction

We are committed to publishing regular updates¹ 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 covers decisions that the board issued during and prior to Q1 2022, and it includes details of (1) Meta's content referrals and Policy Advisory Opinion requests to the board and (2) our progress on implementing the board's non-binding recommendations. The report is meant to strengthen transparency and hold us accountable to the board and to the public.

¹ We base these Quarterly Updates on best practices in human rights reporting principles, corporate disclosures, and goal-tracking reports. These include the Value Reporting Foundation's Integrated Reporting Framework and Sustainability Accounting Standards Board (SASB) Standards, Global Reporting Initiative (GRI) Reporting Principles, and the UN Guiding Principles for Business and Human Rights (UNGPs), among others.

I. Meta's Referrals

In addition to providing people who use Facebook and Instagram with direct access to appeal our content enforcement decisions to the board, we regularly and proactively seek input from the Oversight Board on some of the most significant and difficult content decisions, policies, and enforcement issues through our content referrals and requests for Policy Advisory Opinions (PAOs). We previously outlined how we prioritize cases for Meta content referrals in our <u>Newsroom</u>. The questions posed by both Meta content referrals and PAOs generally involve issues that are severe, large-scale, and/or important for public discourse.

The Meta content referral process begins with an internal review of content decisions that are geographically diverse, cover a wide range of policies found in our Facebook <u>Community</u> <u>Standards</u> and Instagram <u>Community Guidelines</u>, and represent both content we have taken down and content we have left up. Then, teams with expertise on our content policies, enforcement processes, and specific cultural nuances from regions around the world evaluate the candidate cases for significance and difficulty. Finally, we refer the most significant and difficult content decisions to the board. The board has sole discretion to agree or decline to review the content decisions, the board's decisions on Meta content referrals are 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 consider and publicly respond to its recommendations within 60 days. While PAO recommendations are not binding, the board's guidance through the PAO process holds us accountable for our policies, processes, and decisions.

In response to feedback to be more transparent about the substance of our content referrals, we are providing new overviews on the topics of our referrals and will continue to do so in future Quarterly Updates.

Between January 1, 2022 and March 31, 2022, we submitted 4 content referrals and 1 PAO request to the board. The board selected 1 content referral² and 1 PAO request related to Russia's invasion of Ukraine. However, after the board selected the PAO, we <u>withdrew the request</u> due to ongoing safety and security concerns.

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.

- A person on Instagram posted a photo of the dead body of a politician. At the time the content was posted, there was widespread speculation about whether the politician's death was the result of suicide. In the absence of confirmation on cause of death, Meta determined that the content did not violate its Suicide and Self-Injury policy, and allowed the content to remain on the platform behind a warning screen.
- 2. A person on Facebook posted a video depicting an individual using a hammer to damage a Hindu temple as others cheered them on. Meta determined that the content had been shared in a retaliatory context and had been confirmed to be potentially threatening or likely to contribute to a risk of imminent violence by local law enforcement. Accordingly, Meta removed the content as a veiled threat under our Violence and Incitement policy.
- 3. A satirical magazine's Facebook Page posted a cartoon that violated the Child Sexual Exploitation, Abuse and Nudity policy. Meta initially removed the post but later restored it under our newsworthiness allowance. We placed a warning screen on the content and restricted it to adults, ages eighteen and older.
- The Tigray Communications Affairs Bureau posted content calling for violence on its Facebook page. After review by crisis response teams, Meta removed the content for violating our Violence and Incitement policy. The board <u>selected this case</u> on May 10, 2022.
- 5. Meta requested a PAO from the Oversight Board related to Russia's invasion of Ukraine, which we later <u>withdrew</u> due to ongoing safety and security concerns.



⁴When we decide to leave content up – unless there is a newsworthiness or other noted allowance – 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 or Instagram.

⁵ Meta defines "region" according to an analysis of several factors, including the location of the person posting, the language(s) the content includes, and countries/regions referenced in the content.

Progress on Non-Binding Recommendations

As this and previous updates highlight, the board's impact extends far beyond its binding decisions to uphold or overturn our content moderation decisions. Through its recommendations, the board has pushed us to be more transparent about our policies, products, and enforcement systems and has challenged us to think differently about how we can better serve people and their communities through our platforms.

In Q1 2022, because of the board's recommendations we:

- Initiated two new in-depth policy reviews that will likely end with meetings of the <u>Policy Forum</u>. Through these reviews we will:
 - Consider allowing positive discussion of religious and traditional uses of non-medical drugs in our Restricted Goods and Services policy.
 - Clarify our approach to preventing the functional identification of child victims of sexual violence in our Child Sexual Exploitation, Abuse and Nudity policy.
- Completed <u>research</u> examining how people understand and differentiate between bullying and harassment, and found that a single policy encompassing both is the clearest way to organize our rules for these concepts.
- Launched and completed an experiment exploring the impact of increasing the visibility of the Privacy Violation reporting option, and made the increased visibility a permanent product change.
- Published the Community Standards in Assamese and Farsi, making the Community Standards available in a total of 61 translations.
- Undertook new research projects, both qualitative and quantitative, to better understand how we can incorporate people's voices into our appeals and review processes. We expect this to inform our implementation work for several board recommendations.

We continue to make progress on our implementation of the board's recommendations. In our <u>Q2 and Q3 2021 Quarterly Update</u>, we described three challenges we have experienced in implementing the board's recommendations. In the six months since, we've committed to meaningful improvements to achieve the sustained, structural changes the board recommends.

First, we discussed how the pace and volume of recommendations did not allow us sufficient time to adequately assess and implement the board's guidance. Since then, the Oversight Board bylaw

update extending our required response time from 30 days to 60 days has improved our ability to assess feasibility and provide better informed, more thorough responses. For example, the extended window allowed us to launch, finalize, and incorporate lasting product changes from an experiment prompted by a recommendation in the board's Policy Advisory Opinion on Sharing Private Residential Information.

Second, we described the prioritization processes our Central Integrity team uses to allocate resources and create product development timelines. This "roadmapping" process generally occurs every six months and, as we shared previously, we are better able to integrate the board's long term recommendations as we align implementation work to these teams' roadmaps.

Finally, we shared that limiting communication between the board and Meta to formalized, written exchanges led to information asymmetry. To address this, we have added additional briefings and opportunities for the board to ask real-time questions to Meta subject matter experts. In addition, as we shared in our <u>Q4 2021 Quarterly Update</u>, we now hold quarterly briefings with the board to update them on our work to implement its recommendations. Our goal for these Implementation Working Group meetings is to increase transparency with the board about our progress completing recommendations and to solicit their guidance on various challenges, trade offs, and priorities.

For example, in March 2022, members of the Central Integrity team briefed the Implementation Working Group about the product development cycle for launching more specific user messaging in response to several board recommendations. They shared the design and results of specific experiments, metrics used to decide whether to launch product features, and anticipated next steps. Board members asked both real-time and follow-up questions. Most importantly, they provided feedback on Meta's approach and the degree to which it satisfies the board's recommendations. We recognize the need for as much public transparency as possible around these ongoing engagements with the board. We are exploring ways to share additional information on these and other briefings with the board publicly, while also adhering to legal requirements and our commitments to the safety and privacy of people who use our platforms.

The board has had a meaningful impact on our policies and content moderation systems. As this and other Quarterly Updates make clear though, the work to implement the board's recommendations often takes time because of the complexity and scale associated with changing how we explain and enforce our policies and how we inform users of actions we've taken and what they can do about it. We are committed to this work and welcome continued recommendations from the board. We look forward to sharing our progress in future Quarterly Updates.

1. How to Read This Update

From January 2021 through March 2022, the board issued 109 non-binding recommendations. In our <u>Q4 2021 Quarterly Update</u>, we addressed 47 of these recommendations and indicated we would provide more information on 33 in our next update. In this update, we address those 33 recommendations, along with the 22 new recommendations the board included in the decisions it issued in Q1 2022, for a total of 55 recommendations. We categorize our commitments to the board's recommendations as follows:

- 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 due to, for example, 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 are defined as:

- **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.

The below graph depicts the status of each of the current 55 recommendations:



Out of the 22 new recommendations issued in Q1 2022, we do not have updates for 5. As explained in our 60-day response, we will either take no further action on the recommendation or the recommendation was addressed by work Meta already does.³ We are organizing our substantive updates on the 50 open recommendations into three sections:

- A. **Transparency (20 recommendations):** Helping people understand the rules on Facebook and Instagram, what violates them, and the consequences of violating them.
- B. **Policy (20 recommendations):** Ensuring the Facebook Community Standards and Instagram Community Guidelines are clear and align with our values of expression, authenticity, safety, privacy and dignity.
- C. **Enforcement (10 recommendations):** Improving the quality and efficacy of our content moderation operations at scale.

By structuring our updates this way, we aim to facilitate discussion about the progress made in these areas and improve the navigability of this document. In the body of each section, we provide a general overview of our progress. For further detail and the full text of each recommendation, please refer to the <u>Appendix</u>.

³ This applies to recommendation #2, #5, #11, #15, and #16 in the <u>PAO on Sharing Private Residential Information</u>.



2. Non-Binding Recommendation Implementation Highlights

A. Transparency

We are providing updates for our work on 20 board recommendations that address transparency and accountability.

We want to highlight our progress on:

- 1. Conducting ongoing research to ensure that the people who use our platform feel that Meta has given them a voice and heard their point of view.
 - As part of our commitment to building trust in Meta's integrity processes, we conducted global qualitative research to understand how we can help people feel heard through our appeals and review processes. The research demonstrated that these processes should function as a dialogue that promotes mutual understanding,

helping people understand Meta's decision-making and helping Meta understand people's views. These findings will inform future product design and development.

- Similarly, we conducted experiments that improved the efficacy of our enforcement messaging for Hate Speech and Bullying & Harassment policy violations.
- 2. Translating our Community Standards to new languages. In our <u>Q4 2021 Quarterly Update</u>, we shared that we completed our implementation of the board's recommendation from its <u>decision on Punjabi Concern Over the RSS in India</u>. In its recommendation, the board recommended that we translate our Community Standards into Punjabi, and set a goal of making our Community Standards accessible in all languages widely spoken by people who use Facebook and Instagram. In Q1 2022, we continued this work and published the Community Standards in Assamese and Farsi, making the Community Standards available in a total of 61 translations.

For a comprehensive list of all 20 recommendations in this category, see <u>Appendix A.</u> <u>Transparency</u>.

B. Policy

We are providing updates for our work on 20 board recommendations that address the Facebook Community Standards and Instagram Community Guidelines.

We want to highlight our progress on:

1. Undertaking new policy development. As of Q1 2022, 3 policy recommendations from 3 different cases have gone through (or will likely soon be going through) our <u>Policy Forum</u>. As described previously, we hold Policy Forums to discuss potential changes to our Community Standards, Community Guidelines, Advertising Policies, or Product Policies. At these meetings, subject matter experts propose adding new policies or amending existing ones. Increasingly, those proposals are based on board recommendations. Based on input from outside experts, a team of cross-functional stakeholders within the company weigh the merits of these changes relative to varying perspectives on safety and voice, and the impact of our policies on global communities with different cultural and political contexts. In January 2022, Oversight Board staff attended a Policy Forum for the first time. Similarly, the board and board staff will be welcomed to future Policy Forums, including those planned for later this year to discuss proposals based on board recommendations. These are:

- Approach to Discussion of Non-Medical Drugs: Since our previous Quarterly Update, we have begun outreach to a global array of internal and external subject-matter experts regarding the possibility of allowing for the positive discussion of religious and traditional uses of non-medical drugs in our Restricted Goods & Services policies. We expect to present the proposal at a <u>Policy Forum</u> this year and will invite the board and board staff to attend.
- Approach to Functional Identification of Child Victims of Sexual Violence: We have also begun to scope the policy development work to implement the board's recommendation that we determine when and how to incorporate a prohibition on functional identification of child victims of sexual violence in our Child Sexual Exploitation, Abuse and Nudity policy.
- 2. Completing research that examines how people understand and differentiate between bullying and harassment. We conducted a review of prior academic literature on these topics and identified a lack of consensus over their basic definitions. We also found that definitions for these two concepts often share similarities. Next, we conducted a survey of internet users across ten countries (including 29,827 adults and teens, with consent from their parents) to measure how people think about and apply concepts of bullying and harassment.⁴ Based on this research, we found considerable variation in people's perceived definitions of the two concepts and very little evidence that people make clear and consistent distinctions between them. In fact, we found that these internet users characterize many hypothetical situations as both bullying *and* harassment. As a result, we believe that a single policy encompassing both is the clearest way to organize our rules for these concepts. Full details of this research can be found here.

For a comprehensive list of all 20 recommendations in this category, see <u>Appendix B. Policy</u>.

⁴ The ten countries surveyed were Brazil, Egypt, France, India, Japan, Nigeria, Norway, the Philippines, the United Kingdom, and the United States.

C. Enforcement

We are providing updates for our work on 10 board recommendations that address our enforcement systems.

We want to highlight our progress on:

1. Launching and completing an experiment exploring the impact of increasing the visibility of the Privacy Violation reporting option. In response to a board recommendation, we conducted an experiment on increasing the visibility of the option to report Privacy Violations on Facebook. The results showed an increase in Privacy Violation reports, so we have rolled out this update to all people on Facebook. The option to select "Privacy Violation" is now third on the list of suggested report types. We plan to continue this progress by launching a second experiment, examining the impact of adjusting the language of privacy violation reporting to be more specific.

For a comprehensive list of all 10 recommendations in this category, see <u>Appendix C</u>. <u>Enforcement</u>.

II. Appendix

How to Read This Appendix

The board recommendations in this appendix are categorized by implementation commitment level and current status of implementation, as detailed in page 9 under <u>How to Read This Update</u>. For each recommendation, we include the following information:

- **Oversight Board Recommendation**: The board recommendation we are sharing a progress update on, along with any accompanying recommendations. As noted previously, when the board has issued similar recommendations across multiple cases, we combine these recommendations in progress updates.
- **Previous Category**: The implementation commitment level indicated in the last Quarterly Update, or the 60-day response to the board, whichever was more recent.
- **Updated Category**: The current implementation commitment level based on updated assessments or additional input from the board in the form of relevant recommendations, content decisions, or guidance from the Implementation Working Group.
- Current Status: The current status of our implementation work.
- May 2022 Update: An overview of our recent progress, challenges, considerations, and next steps for our work for each recommendation

Appendix A. Transparency

Transparency Recommendations

Oversight Board Recommendation: 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 <u>Breast Cancer Symptoms and Nudity</u> <u>Recommendation #3, Nazi Quote Recommendation #1, Depiction of Zwarte Piet Recommendation</u> #2, South Africa Slur Recommendation #1, Post Discussing a Substance with Psychoactive Properties <u>Recommendation #2, and PAO on Sharing Private Residential Information #17</u>)⁵)

Previous Category	Implementing in Part
Frevious category	
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	We continue making progress on this recommendation as part of our ongoing commitment to build trust in Meta's integrity processes. We completed two experiments relevant to this recommendation, which improved the efficacy of our enforcement messaging. One experiment tailored Hate Speech messaging based on the person's violation type and another improved Bullying & Harassment violation messaging. We are also expanding the number of violation types for which we communicate the specific policy that was violated ir enforcement messaging. We expect to continue testing and improving on these experiences Similarly, we conducted global qualitative research to understand how we might ensure that when people appeal Meta's decision to take down a piece of content that they posted, they feel that Meta has given them a voice and heard their point of view. The research demonstrated that, to ensure people feel heard, the review process needs to function more like a dialogue that promotes mutual understanding: helping people understand Meta's decision-making and Meta understand people's rationale. These findings will inform the second phase of research: prototyping a new user experience in Q3 2022 and beginning to develop new metrics for measuring voice.
•	Board Recommendation: Inform users when automation is used to take enforcement on against their content, including accessible descriptions of what this means. (<u>Breast Cancer Symptoms & Nudity Recommendation #5</u>)

Previous Category	Implementing in Part
Updated Category	Implementing in Part

⁵ The board issued similar recommendations in the following cases: <u>Breast Cancer Symptoms and Nudity #3</u>, <u>Nazi Quote #1</u>, <u>Depiction</u> of Zwarte Piet #2, <u>South Africa Slur #1</u>, <u>Post Discussing a Substance with Psychoactive Properties #2</u>, and <u>PAO on Sharing Private</u> <u>Residential Information #17</u>. 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.

Current Status	In Progress
May 2022 Update	Following an experiment in Q4 2021, in which we informed people whether automation or human review led to their content being taken down, we will launch this messaging in select parts of the world in Q2 2022. We will use data from this pilot launch to further understand the impact of the message on people's experiences, and improve our design and approach globally. We will continue to report on our progress in the next Quarterly Update.

Oversight Board Recommendation : 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	
May 2022 Update	While this work is actively underway, we have not made significant progress since our Q4 2021 Quarterly Update, in which we explained that we are exploring ways to expand our transparency reporting to include more data on our automated enforcement systems. Due to 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. As shared previously, 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. This work will take a significant amount of time but we are currently aiming to complete this recommendation by Q4 2023. We will provide an update on the status of this recommendation in the next Quarterly Update.	

Oversight Board Recommendation : Facebook should more clearly explain its newsworthiness allowance.

(Former President Trump's Suspension Recommendation #11 (along with <u>Post Depicting Protests in Colombia While</u> <u>Using a Slur Recommendation #2</u>)⁶)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	To ensure that people are able to easily reference information on how we make decisions related to newsworthiness, we updated the introduction of the <u>Community Standards</u> in our Transparency Center to link to more information regarding our approach. We are continuing

⁶ The board issued a similar recommendation in recommendation #2 in the <u>Post Depicting Protests in Colombia While Using a Slur</u> case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #11 in the Former President Trump's Suspension case.

18

to identify examples of newsworthiness applications and expect to publish these examples and additional information on our approach to newsworthiness later this year.

Oversight Board Recommendation : 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	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	As we shared in our Q4 2021 Quarterly Update, we are actively working on two long-term initiatives resulting from this recommendation: measuring our enforcement actions on profile, page, and account restrictions; and measuring enforcement data by location. Both of these initiatives fit into our overall vision for the Community Standards Enforcement Report (CSER): that it should accurately represent our evolving enforcement practices. To address the first initiative, we are in the process of improving the granularity of our reporting by beginning to define these metrics and build reliable processes for measuring enforcement of complex entities like Accounts, Groups, and Pages. We are continuing to define metrics related to profiles, page, and account restrictions and to create processes for measuring them accurately before we report them publicly. This work will take a significant amount of time, but we are currently aiming to complete this work by Q4 2023. We will provide an update on the status of this recommendation in the next Quarterly Update.

Oversight Board Recommendation : 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	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	We are beginning to define accuracy metrics this year and aim to launch these metrics by Q4 2023. This is due to existing priorities on our data reporting roadmaps, such as those outlined in our <u>previous Quarterly Update</u> and in our response to <u>Former President Trump's</u> <u>Suspension Recommendation #18</u> . As we explain in our update to <u>Former President Trump's</u> <u>Suspension Recommendation #18</u> , we are also actively working on the challenges of measuring enforcement data by location. We will provide an update on the status of this recommendation in the next Quarterly Update.

Oversight Board Recommendation : 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		
May 2022 Update	As we shared in our <u>Q4 2021 Quarterly Update</u> , to build new messages to inform people about whether their content has been removed as a result of a government request, we have to develop new tools and processes for reporting this information across the range of government request formats we receive. We expect to begin these initial updates to internal infrastructure this year and anticipate completing full development of the data infrastructure by Q4 2023. Once it is built, this infrastructure will allow us to begin assessing options and scoping work for these new types of user messages. We will provide a timeline of milestones for implementing this recommendation in a future Quarterly Update.		
receives for co (as	Oversight Board Recommendation : 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. (<u>Support of Abdullah Öcalan, Founder of the PKK Recommendation #11</u> (along with <u>AI Jazeera Post</u>		
	on Tensions Between Israel and Palestine #4 ⁷)		
Previous Category	Implementing Fully		
Updated Category	Implementing Fully		
Current Status	In Progress		
May 2022 Update	 Since our <u>Q4 2021 Quarterly Update</u>, we have completed initial scoping on how to measure and calculate the following country-level metrics outlining the number of: 1. Unique government requests to remove content we receive 2. Unique pieces of content covered by these requests 3. Such pieces of content removed under the Community Standards 		

⁷ The board issued a similar recommendation in recommendation #4 in the <u>AI Jazeera Post on Tensions Between Israel and Palestine</u> 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.

Oversight Board Recommendation : 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
May 2022 Update	As shared previously, we have partnered with a third party nonprofit to conduct human rights due diligence in line with this recommendation and in accordance with the UN Guiding Principles on Business and Human Rights (UNGPs). This human rights due diligence began in mid-2021, and is still underway. Consistent with our human rights policy and the UNGPs, we hope to publicly communicate insights from this work by Q3 2022.

Oversight Board Recommendation : 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
May 2022 Update	As we explained in our update on our work on <u>Former President Trump's Suspension</u> <u>Recommendation #11</u> , we have updated the introduction of the <u>Community Standards</u> in our Transparency Center to link to more information about our approach to newsworthiness. We are continuing to evaluate ways to inform people when content assessed as violating is left on our platforms due to a newsworthiness allowance. In line with our findings, we aim to implement some user notifications by the end of the year. We will continue to report on our progress in the next Quarterly Update.

Oversight Board Recommendation : 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

May 2022 Update	As shared in our <u>previous Quarterly Update</u> , we plan to update the messaging that people receive when we notify them of a change to the status of their content to explain that the change is a result of the person's appeal to the board. While not finalized, the messaging is likely to 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 hoped to implement this change in Q1, but due to unexpected work based on urgent regulatory priorities, we have had to deprioritize this recommendation for this quarter. However, it will be incorporated into roadmaps for Q3 and Q4 2022. We will continue to report on our progress in the next Quarterly Update.
assessment of rumors that I measures Me should also re reporting of capabilities should co	Board Recommendation : Commission an independent human rights due diligence on how Facebook and Instagram have been used to spread hate speech and unverified heighten the risk of violence in Ethiopia. The assessment should review the success of ta took to prevent the misuse of its products and services in Ethiopia. The assessment eview the success of measures Meta took to allow for corroborated and public interest in human rights atrocities in Ethiopia. The assessment should review Meta's language in Ethiopia and if they are adequate to protect the rights of its users. The assessment over a period from June 1, 2020, to the present. The company should complete the nt 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	Implementing in Part
Current Status	Complete
May 2022 Update	The global Human Rights team has worked as an integral part of the At Risk Countries (ARC) program since 2019. An overview of our approach to countries at risk can be <u>found in our</u> <u>Transparency Center</u> . The team has embedded standards and principles from the UN Guiding Principles on Business and Human Rights (UNGPs) into relevant prioritization and analytical frameworks, which serve as ongoing due diligence tools. Shortly after our global Human Rights team was created, Ethiopia's risk level was re-evaluated, a third party fact checking assessment program was established, and other colleagues were trained on atrocity crime risks and frameworks. Human Rights staff have also participated in Ethiopia risk management processes on an ongoing basis since 2019.
	Since early 2020, due diligence has included field-based information gathering and research, focused stakeholder engagement, and formal due diligence from the Business and Human Rights Group in partnership with our Africa Public Policy team. An updated Human Rights

22

Impact Assessment (HRIA) focused on risk mitigation ahead of Ethiopia's elections was finalized in June 2021.

In 2022, members of the cross-functional team of Meta staff visited Addis Ababa for three days to meet with a wide array of stakeholders that included: the Prime Minister's Office, various government agencies, human rights activists, our Trusted Partners, and independent institutions including the Ethiopia Media Authority and the Ethiopia Human Rights Commission. The team has also engaged with new stakeholders about our Community Standards, digital security, safety and privacy standards, connectivity work, and our social impact initiatives.

This extensive and multi-layered due diligence and stakeholder engagement has informed our approach to content moderation and product development for the region. Our human rights teams continue to support Content Policy teams to understand the humanitarian situation, codewords, and slurs being used on the ground. Relevant members of our Content Policy teams have also received law of armed conflict training. We can now review content in the top four languages spoken and those central to the conflict (Amharic, Oromo, Somali, Tigrinya). Additionally, we have created solutions that make it easier for Ethiopians, as well as specialized international and local human rights and civil society organizations, to notify us when they see potentially violating content, so we can investigate it for possible violations.

As stated previously in response to the board's recommendation, we will continue to conduct thorough human rights due diligence and dynamic risk management processes in Ethiopia, and share insights and actions from our due diligence that aligns with the board's recommendations, our Human Rights Policy, and the UNGPs.

While privacy and safety risks preclude the full publication of our Ethiopia due diligence, we are seeking to release our first ever annual human rights disclosure report in mid 2022. The report will include our approach to due diligence and insights and actions from our ongoing efforts to identify, prevent, and mitigate human rights risks around the world.

Appendix B. Policy

Policy Clarity & Accessibility Recommendations

Oversight Board Recommendation: 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.

(<u>Breast Cancer Symptoms & Nudity Recommendation #2</u> (along with <u>Support of Abdullah Öcalan</u> <u>Recommendation #10, Post Discussing a Substance with Psychoactive Properties</u> <u>Recommendation #1, and PAO on Sharing Residential Information #9)⁸</u>)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our response to Post Discussing a Substance with Psychoactive Properties Recommendation #1, we committed to publishing updates to the Instagram Community Guidelines to reflect the Facebook Community Standards in all shared policy areas; and to make the differences clear in the small number of instances where the policies differ. We agree with the board that we should be clear about the rules that we use to protect people on our platforms. Our work to implement this recommendation has taken longer than anticipated. Since our last update, our company rebranded to Meta to bring our apps and technologies together under one new corporate brand. Our long term ambition is for all of our products to co-exist within the Metaverse with one set of shared values. We believe that the people who use Instagram and Facebook will be best served by unified Community Standards, which clearly explain the differences in our policies between the two platforms. We are working with our legal, regulatory, and product teams to scope and implement this plan, adjusted to reflect our new corporate brand and mission, while still fully implementing the spirit of the board's recommendations. We expect to complete this recommendation by the end of the year.
	ard Recommendation: Facebook should be clear in its Corporate Human Rights Policy preserves, and shares information related to investigations and potential prosecutions, including how researchers can access that information.
	(<u>Former President Trump's Suspension Recommendation #15</u>)
Previous Category	Implementing in Part
Updated Category	No Further Action

⁸ The board issued a similar recommendation in recommendation #10 in the <u>Support of Abdullah Öcalan</u> case, recommendation #1 in the <u>Post Discussing a Substance with Psychoactive Properties</u> case, and recommendation #9 in the <u>PAO on Sharing Residential</u> <u>Information</u> 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.

Current Status	No Further Updates
May 2022 Update	Consistent with Meta's <u>Corporate Human Rights Policy</u> , we continue to explore additional avenues to support international accountability processes—including through the potential identification, preservation, and disclosure of relevant content—within the constraints of our legal and privacy obligations. While we will have no further updates specific to this recommendation, we look forward to sharing additional information on our approach to this important issue in <u>future human rights reporting</u> .

Oversight Board Recommendation: 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
May 2022 Update	Earlier this year, we presented and adopted the Crisis Policy Protocol at the Policy Forum with Oversight Board staff in attendance. Over the past few months, we have been working to build the internal architecture necessary to be able to launch this as planned at the end of Q2 2022. We will provide additional information in a future Quarterly Update.

Oversight Board Recommendation: 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	No Further Action
Current Status	No Further Updates
May 2022 Update	To address the first part of the board's recommendation, we conducted research to evaluate whether people differentiate between bullying and harassment with sufficient specificity to justify separate and distinct definitions and policy treatment. Our review of prior academic literature found a lack of consensus over basic definitions of the two terms, and frequent overlap. We also surveyed 29,827 internet users in 10 countries and found a high level of variation about how people think about and apply concepts of bullying and harassment across countries, people, and hypothetical situations of online behavior. Our research also found that internet users characterize many hypothetical situations as both bullying and harassment. During the survey, participants did not identify any factors in hypothetical situations that strongly push them to apply one label over the other. We were able to conclude there was not sufficient evidence to support separating the definitions of bullying

and harassment in our policies. Therefore, we will have no further updates on this recommendation. You can read more about our research <u>here</u>.

Oversight Board Recommendation: Clearly define Meta's 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 (along with January 2021 Protests in Russia Recommendation #4)⁹)

Previous Category	Implementing Fully
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	As shared in our <u>Q4 2021 Quarterly Update</u> , we updated the Bullying and Harassment section of our Community Standards to clarify our policy based on this recommendation. We shared these changes as part of our August 2021 Community Standards updates, which included the addition of tiers and language clarifications. Rather than providing illustrative examples of each target category, which could allow violators to evade detection and increase the risk of jeopardizing people's safety, we plan to address the spirit of the recommendation by including further details about each target category in our <u>Transparency</u> <u>Center</u> . We are continuing to explore ways to share more details about our Bullying and Harassment policy that align with the structure and readability of the Community Standards, and we aim to share these examples by Q3 2021.

Oversight Board Recommendation: 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
May 2022 Update	In consultation with a number of teams at Meta — including News Partnerships, Human Rights policy, and teams with local context — we are working on developing criteria intended to increase understanding of what constitutes "news reporting" in our <u>Dangerous Individuals</u> <u>and Organizations</u> policy. We are also undergoing a policy review to increase clarity and transparency around our definitions of Praise, Substantive Support, and Representation (PSR) and the allowable carve-outs from those terms, including news reporting and neutral discussions. We expect to conclude this process by Q4 2022. We will also continue to work

⁹ The board issued a similar recommendation in recommendation #4 in the <u>January 2021 Protests in Russia</u> case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #3 in the January 2021 Protests in Russia case.

on ways to share more about these Dangerous Individuals and Organizations policy nuances externally.

Oversight Board Recommendation: Publish illustrative examples from the list of slurs Meta 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	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	We are in the final stages of adding language to our Hate Speech policy that clarifies our approach to identifying slurs that may be used to attack people based on their protected characteristics. We anticipate publishing this updated language in the coming months. We are also continuing to explore ways to include more illustrative examples of the types of terms that qualify as slurs or insulting labels under our policy. As explained in our <u>Q4 2021</u> <u>Quarterly Update</u> , we decided not to publish a comprehensive list of violating slurs, as these terms may create an environment of intimidation and exclusion. Given the sensitive nature of some of these words, we are continuing to carefully consider which examples to include, and how, in order to reduce the potential for harm to anyone reading our Community Standards. We will continue to report on our progress in the next Quarterly Update.

Oversight Board Recommendation: 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
May 2022 Update	We are conducting a policy development process to explore potential changes to our <u>Restricted Goods and Services</u> policy ¹⁰ regarding discussion of the use of non-medical drugs in traditional or religious contexts. As part of this work, we have begun our internal working group process to discuss whether we should make such a change and, if so, what options we should consider. These working groups have included teams from multiple regions and include representation from policy, operations, and research teams, among others. We are also informing any potential changes to the policy by conducting independent research and

¹⁰ In November 2021, our Regulated Goods policy was renamed Restricted Goods and Services

connecting with an array of global external experts in the space. Ultimately, we hope to present a recommendation for this at a <u>Policy Forum</u> this year. We will continue to report on our progress in the next Quarterly Update.

Oversight Board Recommendation: Meta should publish its internal definitions for "non-medical drugs" and "pharmaceutical drugs" in the Facebook Community Standard on Restricted Goods and Services. The published definitions should: (a) make clear that certain substances may fall under either "non-medical drugs" or "pharmaceutical drugs" and (b) explain the circumstances under which a substance would fall into each of these categories. The Board will consider this recommendation implemented when these changes are made in the Community Standard.

(<u>Post Requesting Advice on Pharmaceutical Drugs #1</u>)	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our April 1, 2022 response, we made the commitment to share definitions for both "non-medical drugs" and "pharmaceutical drugs" in our <u>Restricted Goods and Services</u> policy. These definitions will clarify the existing policy distinctions between non-medical drugs and pharmaceutical drugs. We anticipate publishing these definitions in our Transparency Center by Q3 2022.

Oversight Board Recommendation: Meta should define graphic depiction and sexualization in the Child Sexual Exploitation, Nudity and Abuse Community Standard. Meta should make clear that not all explicit language constitutes graphic depiction or sexualization and explain the difference between legal, clinical or medical terms and graphic content. Meta should also provide a clarification for distinguishing child sexual exploitation and reporting on child sexual exploitation. The Board will consider the recommendation implemented when language defining key terms and the distinction has been added to the Community Standard.

(Post Describing Sexual Violence Against Minors #1)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	We are in the initial stages of developing definitions for graphic depiction and sexualization, and expect to publish them in the <u>Child Sexual Exploitation</u> , <u>Nudity and Abuse</u> policy of the Community Standards sometime this year. The definition will capture the distinction between violating content and non-violating content, specifically highlighting that descriptions in legal, clinical, or medical contexts are non-violating. We will also develop and publish clarifying guidelines for "depiction" and "reporting" of child sexual exploitation.

Oversight Board Recommendation: Meta should undergo a policy development process, including as a discussion in the Policy Forum, to determine whether and how to incorporate a prohibition on

functional identification of child victims of sexual violence in its Community Standards. This process should include stakeholder and expert engagement on functional identification and the rights of the child. The Board will consider this recommendation implemented when Meta publishes the minutes of the Product Policy Forum where this is discussed.

(Post Describing Sexual Violence Against Minors #2)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our April 1, 2022 response, we shared that our teams are currently in the initial scoping stages for developing a potential policy that prohibits the functional identification of child victims of sexual violence, with the goal of ultimately bringing options for implementing this change to the Community Standards to <u>the Policy Forum</u> . Consistent with past meetings of the Policy Forum, we intend to inform our policy development with input from a range of internal and external perspectives. This may include conducting new research, consulting internal and external experts in areas such as freedom of expression and safety, and working with teams throughout Meta to understand the feasibility of implementing options. Following a discussion at the Policy Forum, we will post a summary of the proceedings in our <u>Transparency Center</u> . We will continue to report on our progress in the next Quarterly Update.

Oversight Board Recommendation: Meta should remove the exception that allows the sharing of private residential information (both images that currently fulfill the Privacy Violations policy's criteria for takedown and 10 addresses) when considered "publicly available". This means Meta would no longer allow otherwise violating content on Facebook and on Instagram if "published by at least five news outlets" or if it contains residential addresses or imagery from financial records or statements of an organization, court records, professional and business licenses, sex offender registries or press releases from government agencies, or law enforcement. The Board will consider this implemented when Meta modifies its Internal Implementation Standards and its content policies.

(PAO on Sharing Residential Information #1 (along with PAO on Sharing Residential Information #3)¹¹)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our April 8, 2022 response, we committed to fully implementing this recommendation by removing the existing "publicly available" exception to the Privacy Violations policy. In doing so, we will also allow the sharing of imagery that displays the external view of private residences in various scenarios, but not when there is a context of organizing protests against the resident. We are still in the initial scoping stages of this policy development

¹¹ The board issued a similar recommendation in recommendation #3 in the <u>PAO on Sharing Residential Information case</u>. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #1 in the PAO on Sharing Residential Information

process and anticipate implementing this recommendation by the end of this year. We will continue to report our progress in the next Quarterly Update.

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

(PAO on Sharing Residential Information #4)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our April 8, 2022 response, we stated that on Facebook and Instagram we will allow for the organization of protests at publicly owned official residences, in cases where we can accurately identify these locations. In line with the board's guidance, the implementation of this recommendation will not include private residences of government officials. As explained in our response, forming a consistent definition of "high ranking officials" and applying it in countries and communities around the world is a complex undertaking, particularly for local governments. We are still in the initial scoping stages of this work, but we plan on modifying our internal guidance and updating our training materials for reviewers, including using existing guidance from other policy areas to determine whether homes are "publicly owned official residences" for consistency. We will provide an update on the status of this recommendation in the next Quarterly Update.
Violations pol. be removed, specifically ref	oard Recommendation: Meta should better explain, in the text of Facebook's Privacy licy, when disclosing the city where a residence is located will suffice for the content to and when disclosing its neighborhood would be required for the same matter (e.g., by ferencing the population threshold at which sharing only the city as part of the content be considered violating). The Board will consider this implemented when Meta modifies its content policies. (PAO on Sharing Residential Information #7)
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
May 2022 Update	As we outlined in our April 8, 2022 response, we are exploring ways to clarify when

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
May 2022 Update	As we outlined in our April 8, 2022 response, we are exploring ways to clarify when identification of a city—versus identification of a neighborhood or other similar identifier—is sufficient to warrant removal of imagery that displays the external view of private residences. The wide variety of locality types around the world makes it challenging to implement a globally applicable framework, so we will need time to assess what guidance is suitable for content reviewers at scale. While we are still in the initial scoping stages, we hope

	to complete our assessment by the end of the year and will report on our progress in the next Quarterly Update.
•	ard Recommendation: Meta should explain, in the text of Facebook's Privacy Violations ria for assessing whether the resident is sufficiently identified in the content. The Board will consider this implemented when Meta modifies its content policies. (PAO on Sharing Residential Information #8)
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
May 2022 Update	In our April 8, 2022 response, we committed to developing internal criteria for our content reviewers to assess if a resident has been sufficiently identified with a goal to protect privacy and prevent potential harm. To shed more light on our enforcement protocols, we are also working to create additional resources on this issue for our content moderators and published in our Transparency Center. We are still in the initial scoping stages of implementing this recommendation and will report on our progress in the next Quarterly Update.

Appendix C. Enforcement

Enforcement Recommendations

Oversight Board Recommendation: 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, (along with <u>PAO on Sharing</u> <u>Residential Information #14</u>)¹²)

	·
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
May 2022 Update	We agree with the board that we should create more tools for people to provide additional context in their appeals. For this type of functionality, there are informational and other complexities which we are actively considering. For instance, we are aware that many people may not be familiar with the specific exceptions to the Hate Speech policy. In our experience, adding additional steps or options, such as the option to indicate that content falls into a policy exception, can often lead to confusion and discourage people from appealing. We're exploring the right way to offer this function and ensure people have the information they need to use it intentionally, without discouraging appeals or creating unnecessary confusion. We will provide an update on the status of this recommendation in the next Quarterly Update.
Oversight Board Recommendation : 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
May 2022 Update	In our 60-day response to <u>Post Requesting Advice on Pharmaceutical Drugs #2</u> , we committed to assessing the dynamic prioritization of appeals as part of our roadmap planning for Q3 and Q4 2022. As we explained in our response, we generally review appeals in the order we receive them. The only current exception is reviewing people's appeals for their own content we've removed for violating our policies (user appeals) before appeals of decisions when someone reports another person's content for potentially violating our policies (reporter appeals).

¹² The board issued a similar recommendation in <u>PAO on Sharing Private Residential Information #14</u>. We are tracking the progress of our work in response to this recommendations as part of our response to recommendation #4 in the Armenian People and the Armenian Genocide case.

As a part of our assessment, we need to understand the tradeoffs of prioritizing certain appeals over others. There are multiple factors to consider as part of prioritization including, among other things, speed (first come, first serve), severity of enforcement action (an account restriction prioritized over other appeals), and user history (first-time violations get prioritized over potential recidivists). Our engineering, policy, and operations teams are working to better understand these considerations and scope the subsequent product development. We will provide an update on this effort in the next Quarterly Update.

Oversight Board Recommendation : 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.

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	As we shared in our Q4 2021 Quarterly Update, our automated and manual review systems do not currently identify reasons why a piece of content is left up as non-violating. We previously described our effort to assess whether it was feasible to run an analysis on the basis of the limited data that is currently available. Ultimately, we determined that a more system-level option for better understanding the accuracy rates of how we apply policy allowances would be more accurate. Our goal is that this approach will provide the necessary data on both volume of non-violating content and capture labels for which Hate Speech allowances were applied. We expect to complete this accuracy assessment by the end of 2022 and will provide an update in the next Quarterly Update.

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

Oversight Board Recommendation : Meta should study the consequences and trade-offs of implementing a dynamic prioritization system that orders appeals for human review, and consider whether the fact that an enforcement decision resulted in an account restriction should be a criterion within this system. The Board will consider this recommendation implemented when Meta shares the results of these investigations with the Board and in its quarterly Board transparency report.

(Post Requesting Advice on Pharmaceutical Drugs #2)

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

Current Status

Complete

May 2022 Update	In our April 1, 2022 response, we committed to assessing the dynamic prioritization of appeals as part of our roadmap planning for Q3 and Q4 2022. As we explained in our response, we generally review appeals in the order we receive them. The only current exception is reviewing appeals regarding decisions made about a person's own content (user appeals) before appeals of decisions made when someone reports another person's content for potentially violating our policies (reporter appeals). We are currently partnering with our engineering teams to further understand and scope the necessary changes to potentially implement a dynamic prioritization framework within appeals. We will provide an update on the status of this recommendation in the next Quarterly Update.
rates fo recommendat including developmer	rd Recommendation : Meta should conduct regular assessments on reviewer accuracy cused on the Restricted Goods and Services policy. The Board will consider this tion implemented when Meta shares the results of these assessments with the Board, how these results will inform improvements to enforcement operations and policy and, and summarize the results in its quarterly Board transparency reports. Meta may se assessments should be extended to reviewer accuracy rates under other Community Standards. (Post Requesting Advice on Pharmaceutical Drugs #3)
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
May 2022 Update	As we explained in our April 1, 2022 response, we already collect and assess data on the basis of takedowns and restoration — including takedowns under our <u>Restricted Goods and</u> <u>Services policy</u> — to develop our classifiers and review protocols and policies. As part of our commitment to transparency, we will continue to explore ways of adding appropriate accuracy metrics to the <u>Community Standards Enforcement Report</u> for reporting consistent, comprehensive and accurate data. Based on external stakeholder input and regular feedback (including from the board), we believe that consistent expansion of the amount and type of data we publish is the best way to provide meaningful insight into our policies and enforcement efforts. We will provide additional information as we complete these ongoing assessments, including as they relate to our Restricted Goods and Services Policy.
the removal o demonstra	ard Recommendation: Users should have a quick and effective mechanism to request f private information posted by others. We will consider this implemented when Meta ates in its transparency reports that user requests to remove their information are and promptly actioned. This recommendation is not applicable to official residences of high-ranking government officials. (PAO on Sharing Private Residential Information #6)
Previous Category	Assessing Feasibility
Updated Category	Implementing in Part

May 2022 UpdateFollowing our initial April 8, 2022 response, we started an experiment to increase the
visibility of the Privacy Violations reporting option. The experiment ran initially for a sample
of people on Facebook, and it showed an overall increase in Privacy Violation reports.
Because of these positive results, we have rolled this update out to all users. The option to
select "Privacy Violation" is now third on the list of suggested violation types, after
"Intellectual Property" and "Fraud or Scam." In addition, we plan to launch an experiment
that tests new language to replace the current Privacy Violation language, explicitly
referencing "Exposing Personal Info" as the violation. The goal of this second experiment is
to understand whether the additional granularity about the specific type of privacy violation
increases the number of reports and people's satisfaction with the support they receive. If
the results of this experiment are positive, we will implement this change for all people on
Facebook.

This recommendation has been categorized as "Implementing in Part" because of the board's indication that it will consider this recommendation complete once data on the consistency and speed of our response to Privacy Violations are included in public reporting. For an update on our approach to consistent, clear, and accurate transparency reporting, please see our response to Former President Trump's Suspension Recommendation #18.

Oversight Board Recommendation: Meta should let users reporting content that may violate the Privacy Violations policy provide additional context about their claim. The Board will consider this implemented when Meta publishes information about its appeal processes that demonstrate users may provide this context in appeals.

(PAO on Sharing Residential Information #10)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
May 2022 Update	As we shared in our April 8, 2022 response, we are coordinating internally to prioritize the design of an experiment that would provide people with the opportunity to share additional context in the course of an appeal. Because the next steps for this recommendation relate to our plans to address <u>Armenian People and the Armenian Genocide recommendation #4</u> , we will track future updates under that recommendation.

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

(PAO on Sharing Residential Information #12)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility

Current Status	In Progress
May 2022 Update	As we discussed in our April 8, 2022 response, we are exploring ways to incorporate elements of this recommendation into how we enforce violations of our Privacy Violations policy. When someone violates the Community Standards, <u>we sometimes apply a strike to the account</u> . Whether we apply a strike depends on the severity of the content, the context in which it was shared and when it was posted. To assess the feasibility of the board's recommendation in this case, we first need to determine whether it is possible to accurately identify at scale when sharing private residential information "is clearly related to malicious action that created a risk of violence or harassment," and if so, how we can log this type of information. We are still in the initial scoping stages for implementing this recommendation, and will provide an update on its status in the next Quarterly Update.
information w Board will processes tha	ard Recommendation: Meta should give users an opportunity to remove or edit private ithin their content following a removal for violation of the Privacy Violations policy. The consider this implemented when Meta publishes information about its enforcement t demonstrates users are notified of specific policy violations when content is removed nd granted a remedial window before the content is permanently deleted. (PAO on Sharing Residential Information #13)
Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
May 2022 Update	Since our April 8, 2022 response, we are still in the initial scoping stages for designing an early warning system to notify people when content they intend to post may violate our policies. We believe that this system will achieve the spirit of this recommendation by providing people the opportunity to edit or not post potentially violating content. In line with this recommendation, we also plan to launch an experiment exploring the value of tools that would allow people to edit a violating post after Meta removes it in order to avoid a penalty against their account. This experiment will explore giving people the opportunity to edit or remove the parts of their post that violated our Community Standards. We will continue to report on our progress in the next Quarterly Update.

III. Index

Case	Recommendation	Updated Category	Status	Section	Page
Armenians in Azerbaijan	2020-003-FB-UA-1	Implementing in part	In progress	<u>Transparency</u>	16
Breast cancer symptoms and nudity	2020-004-IG-UA-2	Implementing fully	In progress	Policy	23
	2020-004-IG-UA-3	Implementing in part	In progress	<u>Transparency</u> Footnote 5	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	<u>Transparency</u> Footnote 5	16
Former President Trump	2021-001-FB-FBR-11	Implementing fully	In progress	Transparency	17
	2021-001-FB-FBR-15	No further action	No further updates	Policy	23
	2021-001-FB-FBR-18	Implementing in part	In progress	Transparency	18
	2021-001-FB-FBR-19	Implementing fully	In progress	Policy	24
Depiction of Zwarte Piet	2021-002-FB-UA-2	Implementing in part	In progress	<u>Transparency</u> Footnote 5	16
Punjabi Concerns Over the RSS in India	2021-003-FB-UA-3	Implementing in part	In progress	<u>Transparency</u>	18
January 2021 Protests in Russia	2021-004-FB-UA-2	No further action	No further updates	Policy	24
	2021-004-FB-UA-3	Implementing in part	In Progress	Policy	25
	2021-004-FB-UA-4	Implementing in part	In Progress	Policy Footnote 9	25
Armenian People and the Armenian Genocide	2021-005-FB-UA-4	Assessing feasibility	In progress	<u>Enforcement</u>	31
	2021-005-FB-UA-5	Assessing feasibility	In progress	<u>Enforcement</u>	31
Support of Abdullah Ocalan	2021-006-IG-UA-9	Implementing fully	In progress	Transparency	19
	2021-006-IG-UA-10	Implementing fully	In progress	<u>Policy</u> Footnote 8	23
	2021-006-IG-UA-11	Implementing fully	In progress	Transparency	19
Al Jazeera Post on Tensions Between Israel and Palestine	2021-009-FB-UA-1	Assessing feasibility	In progress	Policy	26
	2021-009-FB-UA-3	Implementing fully	In progress	Transparency	20

	2021-009-FB-UA-4	Implementing fully	In progress	Transparency Footnote 7	19
Post Depicting Protests in Colombia While Using a Slur	2021-010-FB-UA-1	Implementing in part	In progress	Policy	26
	2021-010-FB-UA-2	Implementing fully	In progress	<u>Transparency</u> Footnote 6	17
	2021-010-FB-UA-4	Assessing feasibility	In progress	Transparency	20
South Africa Slur	2021-011-FB-UA-1	Implementing in part	In progress	<u>Transparency</u> Footnote 5	16
Depicting Indigenous Artwork and Discussing Residential Schools	2021-012-FB-UA-1	Implementing fully	In progress	Transparency	21
	2021-012-FB-UA-3	Implementing in part	In progress	Enforcement	32
Post Discussing a Substance with Psychoactive Properties	2021-013-IG-UA-1	Implementing fully	In progress	<u>Policy</u> Footnote 8	23
	2021-013-IG-UA-2	Implementing in part	In progress	<u>Transparency</u> Footnote 5	16
	2021-013-IG-UA-3	Assessing feasibility	In progress	Policy	26
Post Discussing the Situation in Ethiopia	2021-014-FB-UA-3	Implementing in part	Complete	Transparency	21
Post Requesting Advice on Pharmaceutical Drugs	2021-015-FB-UA-1	Implementing fully	In Progress	Policy	27
	2021-015-FB-UA-2	Assessing feasibility	In progress	Enforcement	32
	2021-015-FB-UA-3	Implementing in part	In progress	Enforcement	33
Post Describing Sexual Violence	2021-016-FB-FBR-1	Implementing fully	In progress	Policy	27
Sexual Violence Against Minors	2021-016-FB-FBR-2	Implementing fully	In progress	<u>Policy</u>	28
PAO on Sharing	2021-001-FB-PAO-1	Implementing fully	In progress	Policy	28
Private Residential Information	2021-001-FB-PAO-2	No further action	No further updates	Footnote 3	10
	2021-001-FB-PAO-3	Implementing fully	In progress	<u>Policy</u> Footnote 11	28
	2021-001-FB-PAO-4	Implementing fully	In progress	Policy	29
	2021-001-FB-PAO-5	No further action	No further updates	Footnote 3	10
	2021-001-FB-PAO-6	Implementing in part	Complete	Enforcement	33
	2021-001-FB-PAO-7	Assessing feasibility	In progress	Policy	29
	2021-001-FB-PAO-8	Implementing fully	In progress	Policy	30
	2021-001-FB-PAO-9	Implementing fully	In progress	<u>Policy</u> Footnote 8	23

	2021-001-FB-PAO-10	Assessing feasibility	In progress	Enforcement	34
	2021-001-FB-PAO-11	No further action	No further updates	Footnote 3	10
	2021-001-FB-PAO-12	Assessing feasibility	In progress	Enforcement	34
	2021-001-FB-PAO-13	Assessing feasibility	In progress	Enforcement	35
	2021-001-FB-PAO-14	Assessing feasibility	In progress	Enforcement Footnote 12	31
	2021-001-FB-PAO-15	No further action	No further updates	Footnote 3	10
	2021-001-FB-PAO-16	No further action	No further updates	Footnote 3	10
	2021-001-FB-PAO-17	Implementing in part	In progress	<u>Transparency</u> Footnote 5	16