Meta Q3 2022 Quarterly Update on the Oversight Board



TABLE OF CONTENTS

Intr	oduction	3
I.	Meta's Referrals	4
II.	Progress on Oversight Board Recommendations & Institutional Impact	7
	How to Read This Update	10
	Recommendation Implementation Highlights	13
	Transparency	13
	Policy	14
	Enforcement	16
III.	Appendix	17
	How to Read This Appendix	17
	Appendix A. Transparency	18
	Appendix B. Policy	26
	Appendix C. Enforcement	35
IV.	Index	40

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 significant and 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 open recommendations from decisions issued by the board during and prior to Q3 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. This 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 Content Referrals & Requests for Policy Advisory Opinions

There are three ways cases can reach the Oversight Board for review: appeals by people, case referrals by Meta, and requests for Policy Advisory Opinions (PAOs). This means that people who use Facebook and Instagram may appeal our content enforcement decisions directly to the board, and that we also regularly and proactively seek input from the Oversight Board on some of the most significant and difficult content decisions, policies, and enforcement issues we face. We previously outlined how we prioritize cases for Meta content referrals in our Newsroom. Both Meta content referrals and PAOs generally involve issues that are significant, 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 Community Standards and Instagram Community Guidelines, and represent both content we have taken down and content we have left up. 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 challenging of these content decisions to the board. The board has sole discretion to agree or decline to review the content decisions referred through this process. The board's decisions on Meta content referrals and user appeals about Meta's content decisions 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 these recommendations are not binding, the board's guidance through the PAO process holds us publicly accountable for our policies, processes, and decisions. When we receive a recommendation from the board, it is integrated as an additional and important input and we make a substantial effort to consider and implement each recommendation.

Between July 1, 2022 and September 30, 2022, we submitted five content referrals to the board:

- 1. A Facebook page shared a video that showed a former spokesperson of the ruling Indian political party (Bharatiya Janata Party) criticizing the Prophet Muhammad. We took no action on the content as we concluded that it did not violate the Hate Speech policy.
- 2. A former US Congressional candidate posted a video on Facebook during his campaign that called for high severity violence against Democrats. We removed the content for violating the Violence and Incitement policy.
- 3. A news reporting organization in Italy posted content, containing several photos of bare breasts on their Facebook Page, which commented on the UK's decision to ban an advertisement from Adidas. We removed the content for violating the Adult Nudity and Sexual Activity policy because it depicts uncovered female nipples.
- 4. A medical trade union in Sri Lanka posted a link on their Facebook Page and encouraged people to donate pharmaceutical drugs. At the time the content was posted, Sri Lanka was in the midst of a financial crisis, limiting the healthcare system's ability to import pharmaceuticals. Meta allowed the content to remain on the platform.
- 5. A media channel posted a photo and caption on Instagram that provided users with recommendations on "How to Save Yourself (And Others) From A Bad Trip" (referring to non-medical drugs). We removed the content for violating the Restricted Goods and Services policy.

The Oversight Board is currently deliberating on our requests for Policy Advisory Opinions on Meta's cross-check policies and removal of COVID-19 misinformation. We did not send any additional Policy Advisory Opinion requests to the board this quarter.



II. Progress on Oversight Board Recommendations& Institutional Impact

As this update highlights, the Oversight Board continues to serve as a valuable source of external perspective and accountability for Meta. The board's recommendations help improve our approach to content moderation at scale by pushing us to be more thoughtful about the impact of our policies, operations, and products. We respond to every board recommendation publicly and have committed to implement or explore the feasibility of implementing 75 percent of recommendations to date. In its Q2 2022 Transparency Report, the Oversight Board said that our recommendation responses and implementation efforts, "demonstrat[ed] a marked increase in responsiveness from the company." Crucially, the board also shared that it has assessed our responses to its most recent recommendations as almost entirely "comprehensive," stating, "Meta has increasingly been committing to concrete action under a set timeline, and has understood the core concern underlying the Board's recommendations." We believe these improvements are a result of better information-sharing with the board, providing the insight necessary for the board to make more targeted recommendations, and an increasing focus on process improvements across both organizations.

In Q3 2022, because of the board's recommendations we:

- Published <u>our response</u> to the human rights due diligence we commissioned to fulfill the Oversight Board's recommendation that Meta should engage an independent entity to analyze Meta's impact on the Israeli-Palestinian conflict and identify any biases in Meta's moderation of content in Arabic and Hebrew.
- Held a <u>Policy Forum</u> to explore expanding our approach to "positive discussion" of non-medical drugs in traditional or religious contexts and will have more to share on this recommendation regarding entheogen drugs by the end of this year.
- Added <u>definitions</u> of the terms "non-medical drugs" and "pharmaceutical drugs" to our <u>Restricted Goods and Services</u> policy.
- Initiated an in-depth policy review of our approach to preventing the functional identification of victims of sexual violence in our <u>Child Sexual Exploitation</u>, <u>Abuse and Nudity</u> and <u>Adult Sexual Exploitation</u> policies. As with our policy development on Non-Medical Drugs, we plan to use input from external and internal experts as well as research to inform this policy development with a goal of discussing it at a future <u>Policy Forum</u>.

- Continued an in-depth review of our definition of "praise" in the <u>Dangerous Individuals and</u>
 <u>Organizations</u> policy, with the goal of allowing for more user voice without sacrificing
 safety in a way that can be equitably operationalized at-scale.
- Launched new notifications to inform people exactly which Community Standard their content violated. This messaging is available globally in English, and translation work is currently underway to make the messaging available in all languages used on our platforms.

Expanding the Oversight Board's Scope

The positive impact the board's oversight and recommendations have had on our work is evident, and we are committed to continued progress towards improved transparency and global impact. Approximately two years after the Oversight Board's formation, in July 2022, Meta approved a new three-year, \$150 million commitment to fund the Oversight Board and its operations. We provided these funds to the Oversight Board Trust, which enables the board to continue operating as a distinct entity, independent from Meta. As stated in the Oversight Board Charter and Bylaws, funds contributed to the Trust are irrevocable and can only be used to fulfill the Trust's purpose and operational and administrative duties. Meta intends to ensure the board remains funded because its efficacy serves as a crucial input for improving our global impact. Each Oversight Board decision provides industry and regulators with a deeply considered, global, human rights-based framework for similar and related content moderation issues. We are supportive of the steps the board has recently taken to increase its impact. We believe that the additional funding, together with its increasing organizational maturity, will enable the board to increase its efficiency and capacity. In turn, we hope that will provide us, users and the board's global audience with a deeper pool of independent, principled guidance to draw from.

In October, the board <u>announced</u> its seven strategic priorities for Meta's approach to content moderation. These strategic priorities, which include elections and civic space, crisis and conflict situation, and gender, reflect themes from cases submitted to the board since its inception, as well as issues facing users globally. We believe this prioritization reflects the evolution of the Oversight Board as an established body and expect that its focus on these areas will enable the board to issue a higher volume of strategic decisions with greater speed.

Following our progress update in the <u>Q2 2022 Quarterly Update on the Oversight Board</u>, we have expanded the board's decision capacity from a binary decision to overturn or uphold our content moderation decisions. Now, if the board determines that content should be restored or

remain on our platforms, it can also issue a binding judgment about whether that content qualifies for the application or removal of a warning screen. Warning screens can be applied to eligible content that does not violate our Community Standards but may still be disturbing or sensitive to people, in order to protect their underlying expression while allowing individuals to choose whether they want to view the content. We see the use of screens like these as an important tool for nuanced content moderation and a valuable opportunity to seek expanded board guidance.

Sharing Information With the Oversight Board

We recognize that the board's oversight capacity is contingent on our ability to share comprehensive information about our content moderation policies and systems. We have made important strides in our efforts to be transparent with the board. As the board described in its Q2 2022 Transparency Report, we continue to answer the vast majority of its questions with comprehensive responses to help inform its case and PAO deliberations.

This quarter, we held a total of 16 briefings for the board. 12 of these were subject matter briefings, covering topics from our user appeals flows and infrastructure to enforcement of our <u>Dangerous Individuals and Organizations</u> policy. A series of these briefings focused on our company-wide approach to human rights, with sessions covering our first annual human rights report, our ongoing approach to human rights due diligence, and the recently-published third-party human rights due diligence on Meta's impacts in Israel and Palestine. In our Q2 2022 Quarterly Update, we highlighted the launch of our Crisis Policy Protocol, developed in response to a May 2021 <u>recommendation</u> by the board. This quarter, we briefed board members and staff on the mechanics of the framework and how it will allow us to assess and designate crisis categories for situations that may require unique policy responses. We answered the board's questions on protocol implementation, capacity constraints, and integration with ongoing preparedness efforts in countries at risk. We also held four Q&A sessions to help inform the board's deliberations on cases and PAOs. Finally, we invited members of the board's staff to shadow fieldwork for a global research sprint assessing the legitimacy of independent accountability tools for tech governance. This months-long effort offered the board a window into our integrity research methodologies, priorities, and findings.

As mentioned in our <u>previous quarterly update</u>, in addition to these information-sharing initiatives, we hold regular Implementation Working Group sessions for the board and board staff to receive updates on our work to implement the board's recommendations and to discuss the feasibility of those recommendations. It is critical to acknowledge the challenges inherent to some

recommendations and to be transparent about timelines and shifting priorities, especially in the face of resource constraints and growing compliance requirements. This quarter, members of our Community Standards Enforcement Report (CSER) team briefed the board's implementation staff on their roadmapping priorities and approach to board recommendations. The wide-ranging discussion covered the procedural and integrity challenges of public data transparency and the intersections between internal roadmapping and regulatory compliance. Forums like these allow the board to develop strategic recommendations that can steer our ongoing and future internal workstreams. They allow us to solicit board guidance early on in the prioritization of our recommendation implementation efforts to ensure that our approaches are well-calibrated to achieve the board's primary aims.

We are also working to share more quantitative information with the board while preserving our legal, safety, and privacy obligations. Increased data-sharing will bolster the quality of our responses during the board's case and PAO deliberations and demonstrate our implementation of past and current recommendations. It will help us to better gauge the scale of the board's impact on the people who use our platforms. We are continuously working to improve the granularity of our reporting by defining more integrity metrics and building better processes for measuring our board recommendation implementation efforts. This is a priority for us, and we see that focus echoed in the board's recommendations and public priorities.

As the progress reported here demonstrates, the board's global expertise, human rights lens, and independence play a crucial role in our ongoing efforts to integrate external accountability and focus on our long-term impact.

1. How to Read This Update

From January 2021 through September 2022, the board issued 130 non-binding recommendations. In our Q2 2022 Quarterly Update, we addressed 56 of these recommendations and indicated we would provide more information on 45 in our next update. In this update, we address those 45 recommendations, 10 new recommendations the board included in the decisions it issued in Q3 2022, and one recommendation from a previous cycle which we are revisiting for consistency with the board's tally², for a total of 56 recommendations. We categorize our commitments to the board's recommendations as follows:

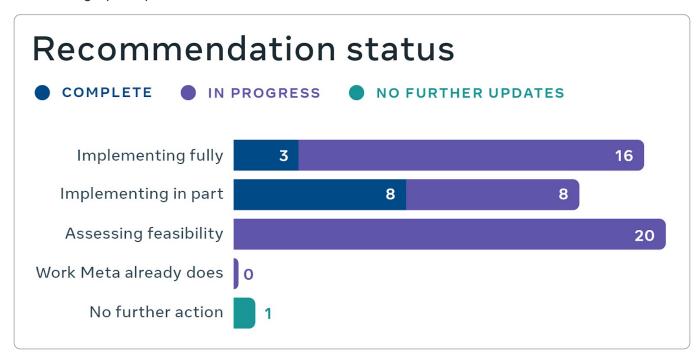
² This applies to <u>Breast Cancer Symptoms & Nudity Recommendation #7</u>, a recommendation from Q1 2021. At the time, Meta and the Oversight Board did not align on recommendation counts, resulting in different recommendation tallies for that case. We hope that, by correcting this misalignment now, it will further our information-sharing goals going forward.

- 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 56 recommendations:

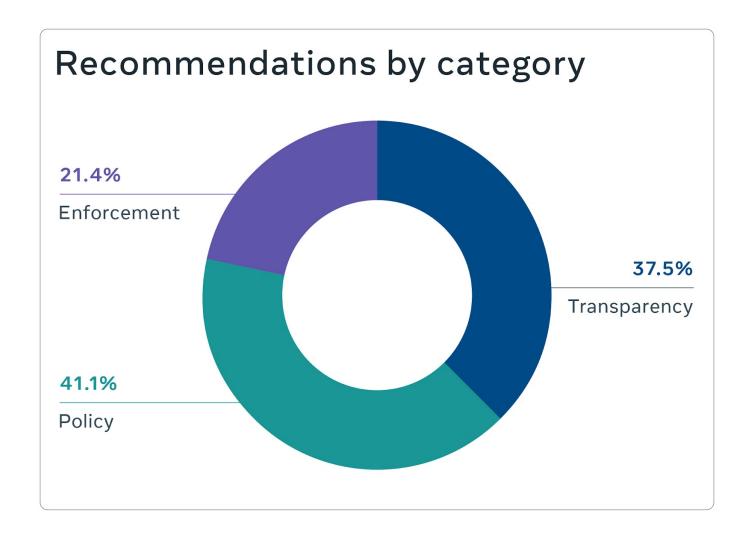


Of the 10 new recommendations issued in Q3 2022, we do not have an update for one. As explained in our 60-day response, we will either take no further action on the recommendation or the recommendation was addressed by work Meta already does.³ We organize our substantive updates on the 55 open recommendations into three sections:

- A. **Transparency (21 recommendations):** Helping people understand the rules on Facebook and Instagram, what violates them, and the consequences of violating them.
- B. **Policy (23 recommendations):** Ensuring the Facebook Community Standards and Instagram Community Guidelines are clear and align with our values of voice, authenticity, safety, privacy, and dignity.
- C. **Enforcement (12 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 Appendix.

³ This applies to recommendation #3 the <u>Cartoon Depicting Violence by the Police in Colombia</u>



2. Oversight Board Recommendation Implementation Highlights

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:

 Completing work on a recommendation that asked us to commission a third-party human rights due diligence report on our impacts in Israel and Palestine: We published our response to the human rights due diligence we commissioned to fulfill the Oversight Board's recommendation that Meta should engage an independent entity to analyze Meta's impact, including any biases in Meta's moderation of content in Arabic and Hebrew. The report identified good practices, areas for improvement, and lessons learned by Meta and made a series of recommendations, consistent with Meta's responsibilities under the UNGPs to take appropriate action to address adverse human rights impacts. Since we received the final report, we've carefully assessed these recommendations to help us learn where and how we can improve. Our response details our commitment to fully implementing 10 of the recommendations, partly implementing four, and assessing the feasibility of another six.

2. Launching new global messaging explaining to people which policy caused us to remove their content: We launched new notifications to tell people exactly which Community Standard their content violated for most violation types. This messaging is available globally in English, and translation work is currently underway to make the messaging available in all languages used on our platforms. We have also updated our enforcement communication framework to reflect both the availability and the transparency level of our communications related to each policy. This updated transparency framework is an effort to comply with board recommendations and emerging content regulations related to the theme of increased granularity in user notifications, and will ensure that we uphold this standard of specificity in all such communications going forward.

For a comprehensive list of all 21 recommendations in this category, see <u>Appendix A.</u> <u>Transparency</u>.

B. Policy

We are providing updates for our work on 23 board recommendations that address the Facebook Community Standards and Instagram Community Guidelines.

We want to highlight our progress on:

 Publishing new definitions for "non-medical drugs" and "pharmaceutical drugs" in our Community Standards: We added definitions of the terms "non-medical drugs" and "pharmaceutical drugs" to our Restricted Goods and Services policy and clarified the circumstances under which a substance might fall into these categories.

- 2. Moving forward with a new approach to allowing positive discussion of religious and traditional uses of non-medical drugs in our Restricted Goods and Services policy following our June Policy Forum: We held a Policy Forum in June to explore expanding our approach to "positive discussion" of non-medical drugs in traditional or religious contexts. We received feedback from a number of different teams and aligned on follow-up action items: conducting legal risk analysis and consulting additional regional teams. After completing these action items, we are now moving forward with the recommendation and expect to share more details by the end of the year.
- 3. Reviewing our approach to Violent and Graphic Content in Human Rights Abuse Contexts: We launched a policy development process on violent and graphic content shared in human rights abuse contexts. Currently, we remove content that is particularly violent or graphic. However, in some contexts, removing this content risks suppressing speech that may raise awareness of human rights abuses. Conversely, allowing it could expose users to disturbing imagery, and risks infringing on the dignity of the people depicted. Therefore, we are examining our approach to violent and graphic content when shared in human rights abuse contexts. As with past policy development processes, we will solicit input from internal and external subject matter experts and conduct comprehensive research to inform our updated approach to the policy.
- 4. Continuing a policy development process to examine our policies and enforcement of "praise" of designated entities on our platforms: We've continued policy development on our definition of "praise" in the Dangerous Organizations and Individuals policy, with the goal of allowing for more user voice without sacrificing safety in a way that can be equitably operationalized at-scale. This process includes soliciting input from external stakeholders from around the world, conducting research to inform this work, and working with our operations teams to capture the complexities of enforcement of "praise" for at-scale reviewers. We expect this work to continue into 2023.

For a comprehensive list of all 21 recommendations in this category, see Appendix B. Policy.

C. Enforcement

We are providing updates for our work on 12 board recommendations that address our enforcement systems.

We want to highlight our progress on:

1. Exploring new improvements to our appeals functionality to allow users to provide more context in their appeals: In our last update, we shared that our product teams were focusing on key drivers of trust in appeals in order to improve our appeals functionality and ensure that people on our platforms feel heard and understand our content moderation decisions. This quarter, our product and content designers have been hard at work on a new design vision for enforcement and appeals experiences. These updates are designed to allow users to provide additional context about why they are appealing an enforcement decision. These product updates include options that would enable users to advocate that their content be allowed to stay on our platforms on the basis of a policy exception. There is still more work that needs to be done to validate these designs before they are potentially implemented.

For a comprehensive list of all 12 recommendations in this category, see <u>Appendix C</u>. Enforcement.

III. 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 10 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.
- **November 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 Recommendation #3, Nazi Quote Recommendation #1, Depiction of Zwarte Piet Recommendation #2, South Africa Slur Recommendation #1, Post Discussing a Substance with Psychoactive Properties Recommendation #2, and PAO on Sharing Private Residential Information #17)⁴)</u>

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
November 2022 Update	In our Q2 2022 Quarterly Update, we announced the global launch of new messaging that notifies people about the specific policy that prompted us to take an enforcement action on their content. This more granular level of detail is an important step forward in transparency and policy education for the people who use our platforms. We have expanded this messaging substantially, and it will apply to the vast majority of violation types on Facebook by the end of the year with expansion to Instagram phased into next year. These notifications are globally available in English and translation work is currently underway to make the messaging available in all languages used on our platforms. We have also updated our enforcement communication measurement framework to reflect both the availability and the transparency level of our communications related to each policy. As described previously, our review systems are often most accurate at the policy level, with the benefit of complete context. Specific, yet inaccurate messaging that lacks context could create worse experiences than correct, broader messaging. With that said, we understand the benefit of even more granular detail, and will continue assessing the feasibility of further increasing the specificity of these notifications. As part of our broader transparency efforts, we will continue to dedicate resources to provide more specific and clear notifications about our enforcement decisions. While we now consider this recommendation complete and will have no further updates on this recommendation, the team understands that this will continue to be a long-term priority as they optimize the system and account for new policy violations in the future.

Oversight Board Recommendation: Inform users when automation is used to take enforcement action against their content, including accessible descriptions of what this means.

⁴ The board issued similar recommendations in the following cases: <u>Breast Cancer Symptoms and Nudity #3, Nazi Quote #1, Depiction of Zwarte Piet #2, South Africa Slur #1, Post Discussing a Substance with Psychoactive Properties #2, and PAO on Sharing Private Residential Information #17. 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.</u>

(Breast Cancer Symptoms & Nudity Recommendation #5)		
Previous Category	Implementing in Part	
Updated Category	Implementing in Part	
Current Status	In Progress	
November 2022 Update	In alignment with our Q2 Quarterly Update, our teams continue to roll out messaging to inform people whether automation or human review led to the removal of their content. Following the pilot of this messaging in France last quarter, our teams will continue launching this messaging in more markets, using findings from these launches to further understand the impact of the message on people's experiences, and improving our design and approach for new markets. As we focus on planning for the upcoming year, this effort aligns with Meta's transparency goals and regulatory compliance requirements and remains a key priority. The additional work required to duplicate the effort for Instagram will be phased into next year and we will provide an update on the status of this implementation in our 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
November 2022 Update	We have made no significant progress since our <u>last quarterly update</u> , where we shared that this work is currently underway with the same long term goal of completion by 2024. As previously stated, we are currently working on what we internally call "data readiness," which requires aligning on a consistent accounting methodology. We are working to define binaries for each metric as a first step towards aggregating public-facing enforcement metrics. To do this, we are discussing complexities such as how to quantify instances of enforcement conducted by human review and automated tools (e.g. quantifying cases where a human reviewer determined that an image was violating and then a machine scaled that decision more broadly). Concurrently, we are resolving gaps in our logging infrastructure to allow us to pull those metrics once we've decided on how to report it. This remains a long-term goal, but we will continue to provide updates on this recommendation in future Quarterly Updates.

Oversight Board Recommendation: Revise the "short" explanation of the Instagram Community Guidelines to clarify that the ban on adult nudity is not absolute.

(Breast Cancer Symptoms & Nudity Recommendation #7)

Previous Category	N/A - Omitted from Tracking

	Assessing Feasibility
Current Status	In Progress
November 2022 Update	This is a recommendation from <u>Q1 2021</u> . At the time, the numbers Meta and the Oversight Board tracked recommendations with different assigned numbers, resulting in recommendation tallies for that case that did not align. We hope that correcting this misalignment now will increase clarity going forward.
	In our 60-day response to <u>Breast Cancer Symptoms & Nudity Recommendation #2</u> in Q1 2021, we explained that we updated the longform Instagram Community Guidelines to read "photos in the context of breastfeeding, birth-giving and after-birth moments, health-related situations (for example, post-mastectomy, breast cancer awareness or gender confirmation surgery) or an act of protest are allowed." In doing so, we considered this recommendation bundled under the assumption that we had addressed it by clarifying that the ban on adult nudity is not absolute. Since then, however, the Oversight Board has alerted us that they track this recommendation separately and that their reporting indicates that we have omitted it from public tracking. To address this discrepancy, we are acknowledging it here. In recent quarters, we have assessed that the best way to clarify the rules that apply to Instagram is to unify the Instagram Community Guidelines, both long- and short-form, and Facebook Community Standards. Because the next steps for this recommendation relate to our plans to address <u>Breast Cancer Symptoms & Nudity Recommendation #2</u> , above, we will track future updates under that recommendation.
_	pard Recommendation: In its transparency reporting, Facebook should include numbers 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	
Previous Category Updated Category	(Former President Trump's Suspension Recommendation #18)
	(Former President Trump's Suspension Recommendation #18) Implementing in Part

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.

(<u>Punjabi Concern Over the RSS in India Recommendation #3</u>)	
Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
November 2022 Update	As mentioned in our Q2 Quarterly Update, our work defining accuracy metrics is underway and we aim to launch these metrics by the end of 2023. In the interim, our teams met with the board to expand on CSER roadmapping practices and provide insight on the process and challenges of public data reporting. Among other issues, the conversation covered the following: Preserving reader accessibility and clarity in increasingly comprehensive reporting Challenges of country level reporting including the base definitions of a country/market An evolving adversarial space where users leverage transparency to propagate platform abuse resulting in challenges We continue to dedicate efforts towards maturing our efforts toward this goal and 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
November 2022 Update	Last quarter, we launched updated user notifications for situations in which we restrict access to content for users in a particular jurisdiction or jurisdictions on the basis of local law following a government request. We detailed that these updated notifications provide consistent messaging across Facebook and Instagram, indicating that content was restricted, and provide a link to a newly-expanded explanation of our process for reviewing government requests within the Content Restrictions section of our Transparency Center. This page also describes how we process these requests and ensure that our actions align with our commitments as a member of the Global Network Initiative and under our Corporate Human Rights Policy. Our goal is to share more detail in user notifications for content restricted based on local laws. We note that we already notify people when their content is removed based on a

violation of our <u>Community Standards</u>, regardless of the way that violation was identified. Work to specifically indicate in this notification when content removed for violating the Community Standards was reported by a government entity identifiable as such is ongoing. Our teams have identified some barriers for disclosure in particular countries that prevent user notice with the same level of transparency due to confidentiality obligations in those jurisdictions. We will provide updates on this work in future Quarterly Updates.

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.

(Support of Abdullah Öcalan, Founder of the PKK Recommendation #11 (along with Al Jazeera Post on Tensions Between Israel and Palestine #4⁵))

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
November 2022 Update	As we shared in our most recent Quarterly Update, we are in the process of developing consistent and reliable systems for gathering metrics on the number of pieces of content removed under the Community Standards as a result of government requests. The objective is to produce government takedown request metrics in the most efficient manner given ongoing challenges including confidentiality obligations and data logging and taxonomy gaps from internal systems. We continue to evaluate approaches to building the necessary internal data logging infrastructure to enable us to publicly report this information across the diversity of request formats in which we receive it. As we shared previously, this remains aligned with Meta's long term integrity goals, but we expect it to be a complex, long-term project. We will provide an update on the timeline for public reporting of these metrics in a future Quarterly Update. In the meantime, we also shared that we have developed our internal policy process for operationalizing sharing information with Lumen — an independent research project hosted by Harvard's Berkman Klein Center for Internet & Society. We remain committed to our public commitment in this regard as we await legal privacy reviews expected to finalize by the end of the year. We will provide an update on the status of this work in our next Quarterly Update.

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

⁵ The board issued a similar recommendation in recommendation #4 in the <u>Al 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.

Updated Category	Implementing Fully
Current Status	Completed
November 2022 Update	On September 22, 2022, our third party partner Business for Social Responsibility (BSR) published its <u>report</u> : Human Rights Due Diligence of Meta's Impacts in Israel and Palestine. We published a <u>response</u> to the report and the 21 recommendations within it. As mentioned in the initial response, we are committed to implementing ten, partly implementing four, and assessing the feasibility of implementing another six of the recommendations. This work will require time as we work with teams internally to address these changes. In the meantime, we briefed and answered the board's questions on the report's methodology and findings. We now consider this recommendation complete and will have no further updates.

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.

(<u>Post Depicting Protests in Colombia While Using a Slur Recommendation #4</u> (along with <u>Video</u>

<u>Depicting a Civilian Victim of Violence in Sudan #4</u>6))

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	As shared in our Q2 2022 Quarterly Update, in addition to updating the introduction of the Community Standards, we added additional detail to our Approach to Newsworthy Content page in the Transparency Center. The new detail includes the annual number of newsworthy allowances issued and sharing examples of newsworthy content. We also continue to evaluate ways to inform people when content assessed as violating is left on our platforms because it is considered newsworthy. We have initiated foundational work on this feature by auditing the current user notification content and increasing the percentage of reports that prompted follow up communications from Facebook. We are also working on increasing the visibility of these communications. This foundational work on post-report communications is intended to increase technical quality and make sure users feel supported by Meta, which will ensure that we are able to implement compliance requirements, as well as this and other recommendations related to specific user messaging in the future. 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

⁶ The board issued a similar recommendation in recommendation #4 in the <u>Video Depicting a Civilian Victim of Violence in Sudan</u> case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #4 in the Post Depicting Protests in Colombia While Using a Slur case.

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)

(Depicting Indigenous Artwork and Discussing Residential Schools Recommendation #1)	
Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
November 2022 Update	In our <u>previous</u> update, we detailed that Meta currently notifies all reporters about its enforcement decision following their reporting of a piece of content. We explained that this includes a secondary notification about further decisions if the person who reported content appeals the result of our decision following their initial report. In the <u>Q2 2022 Quarterly Update</u> we shared that in response to the recommendation, we would launch specific messaging for incorrect outcomes and ensure that it is available in all relevant languages as we develop this solution. While design work is near complete, we have had to prioritize related compliance requirements which has reduced execution capacity on this board recommendation. While the work to launch specific messaging for incorrect outcomes is still an ongoing priority, it will likely not be complete by the end of 2022, as originally scoped. We will share future updates on the related team's strategy and capacity in the interest of ongoing transparency around board priorities.

Oversight Board Recommendation: In line with Meta's commitment following the "Wampum belt" case (2021-012-FB-UA), the Board recommends that Meta notify all users who have reported content when, on subsequent review, it changes its initial determination. Meta should also disclose the results of any experiments assessing the feasibility of introducing this change with the public. The Board will consider this recommendation implemented when Meta shares information regarding relevant experiments and, ultimately, the updated notification with the Board and confirms it is in use in all languages.

(Video of an Edited Cartoon Depicting a Croatian City #2)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
November 2022 Update	We have not made significant progress since our <u>initial response</u> to this recommendation, where we explained that currently when someone appeals a content decision, or reports a piece of content as violating, we notify them when we confirm our original decision to take their content down or decide to restore it. People reporting content who make a secondary appeal to the Oversight Board receive a notification about the board's decision related to their report both in the board's portal and in the Support Centers on Facebook and Instagram. However, the messaging that people receive is standardized and does not specify if a change in Meta's treatment of their content was the result of the Oversight Board's

review process. Consistent with our response to <u>Depicting Indigenous Artwork and Discussing Residential Schools Recommendation #1</u>, we will update the messaging that people receive when we notify them of a change to the status of their content because of an appeal to the Oversight Board. Work to launch this more specific messaging is in progress, and we will ensure that it is available in all relevant languages as we develop and scale the notification.

In addition to undertaking work as a direct result of this recommendation, increasing the specificity of support messaging, including reporter notifications, is a key priority for our product integrity teams. As previously described, we already send initial and secondary notifications to reporters when they report and if/when they appeal an enforcement decision following their report. Going forward, we are working to also share updates on subsequent changes to content status even when a reporter has not appealed our previous decision. These improvements will provide a technical basis for developing the secondary reporter notifications this recommendation calls for. We expect to report progress against this work in 2023.

Oversight Board Recommendation: To enable the establishment of metrics for improvement, Meta should publish the error rates for content mistakenly included in Media Matching Service banks of violating content, broken down by each content policy, in its transparency reporting. This reporting should include information on how content enters the banks and the company's efforts to reduce errors in the process. The Board will consider this recommendation implemented when Meta includes this information in its Community Standards Enforcement Report.

(Cartoon Depicting Violence by the Police in Colombia #3)

Previous Category	No further action
Updated Category	No further action
Current Status	No further update
November 2022 Update	In our <u>initial response</u> to this recommendation, we highlighted our commitment to gathering and sharing accuracy and precision metrics around our content moderation tools, including our media matching systems. This remains a high priority for our Community Standards Enforcement Reporting (CSER) and transparency efforts. However, upon consideration, we determined that sharing metrics for individual MMS banks without sufficient context would likely create confusion, and would not provide a complete picture of the accuracy of our automated enforcement actions. This is because MMS is one component of a larger enforcement system, and media matching does not always work in isolation. While we continue to work towards public reporting of new metrics that will provide comprehensive insights into our enforcement systems, we will have no further updates on this recommendation.

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.

(Breast Cancer Symptoms & Nudity Recommendation #2 (along with Support of Abdullah Öcalan Recommendation #10, Post Discussing a Substance with Psychoactive Properties

Recommendation #1, and PAO on Sharing Residential Information #9)7)

Previous Category	Assessing feasibility
Updated Category	Assessing feasibility
Current Status	In Progress
November 2022 Update	To address the board's recommendations that we improve clarity about which rules apply on Facebook and Instagram, we have built and launched new digestible and user-friendly resources to help educate people about our Community Standards. The first of these resources are three videos coming out of a new program focused on consumer education about integrity topics. The easily digestible and user-friendly videos aim to teach people what our rules are, how and why we update our rules, and how we enforce our rules. The videos live on the Transparency Center and are translated into 61 languages. The goals here are to increase awareness and comprehension of our policies and decrease the number of content violations. We plan to continue this work by making integrity information more approachable with animations, videos, and other short-format media, and by delivering that information to people in contextually relevant situations in-product. We will share additional progress about this work in future Quarterly Updates. Additionally, as described previously, because of our company-wide shift to Meta we are working to unify the Instagram Community Guidelines and Facebook Community Standards. 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 have completed the initial scoping of this recommendation and hoped to launch the unified Community Standards on our new Meta.com domain by the end of the year, but due to unexpected competing product work based on urgent regulatory compliance priorities, we will not complete implementation before the end of the year as anticipated. We will continue to report on our progress in the next Quarterly Update.

⁷ The board issued similar recommendations 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 Information</u> case. We are tracking the progress of our work in response to these recommendation as part of our response to recommendation #2 in the Breast Cancer Symptoms & Nudity case.

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.

(<u>January 2021 Protests in Russia Recommendation #3</u> (along with <u>January 2021 Protests in Russia Recommendation #4</u>)⁸)

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	Complete
November 2022 Update	We previously reformatted our <u>Bullying and Harassment policy</u> in the Community Standards to align with part of this recommendation. In late 2021, we also updated our Community Standards to clarify who we consider a limited scope public figure. As shared <u>previously</u> , we assessed a number of ways to implement the second part of this recommendation without impacting the accessibility of the Community Standards. Ultimately, while our Community Standards aim to provide clarity to what is allowed or not allowed on our platforms, we also do not want to share or promote potentially harmful and negative content in the space designed to prevent that very content. Therefore, we do not plan on adding examples of violating content at this time, and now consider this recommendation implemented in part. We will have no further updates on this recommendation.

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 #19)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	We are undertaking a policy development process to examine potential changes to our existing definition of "praise" in the <u>Dangerous Individuals and Organizations</u> policy. This process is informed by inputs from external experts; academic research; and our internal Operations, Human Rights, News Partnerships, and regional teams. Proposed changes to this policy would also provide further clarity towards enforcement and allowance for neutral discussion, condemnation, and news reporting. Work on this recommendation is still in progress and we will report on our progress in future Quarterly Updates.

⁸ 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.

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
November 2022 Update	In June 2022, we presented a policy recommendation on the topic of "positive" discussion of non-medical drugs in traditional or religious contexts. As part of this policy development, we conducted research and engaged with external experts to consider potential tradeoffs in changing the existing policy restricting discussion of these substances.
	Following a discussion at our Policy Forum and follow-up conversations and considerations, we have aligned on a recommended change to our policy that we intend to implement and publish by the end of the year.

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.

(Post Requesting Advice on Pharmaceutical Drugs #1)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	Complete
November 2022 Update	In line with our <u>Q2 2022 Quarterly Update</u> commitment, we introduced new definitions with more details about what we consider "non-medical drugs" and "pharmaceutical drugs" to our <u>Restricted Goods and Services policy</u> in October. We now consider this recommendation complete and, while we will continue to explore ways to incorporate the board's guidance in this decision into our future work on this type of content, we will have no further updates on this recommendation.

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
November 2022 Update	We are continuing work developing these definitions to clarify "graphic depictions and sexualization" in the <u>Child Sexual Exploitation</u> , <u>Abuse</u> , <u>and Nudity Community Standard</u> . This work includes outlining distinctions between legal, clinical, or medical terms and graphic content. We hope to share more details in the next Quarterly Update.

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
November 2022 Update	We are continuing policy development on functional identification of victims of sexual assault. This work will cover potential changes to the <u>Adult Sexual Exploitation</u> policy as well. In addition to academic and field research, we plan to gather input from internal and external experts. We will share more updates within the next year on the progress and outcomes of this policy development.

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)10)

Previous Category

¹⁰ 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.

Updated Category	Implementing Fully
Current Status	In Progress
November 2022 Update	We are in the process of soliciting final feedback for changes to the <u>Privacy Violations Policy</u> that will remove the "publicly available" allowance for private residential information while still allowing the sharing of imagery that displays the external view of private residences in various scenarios except when organizing protests against the resident. This work will include defining how to identify when depiction is "the focus of the news story" and how to identify protest contexts. We expect to have a more robust update in our 2023 Q1 or Q2 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
November 2022 Update	We are finalizing guidance and training materials for identifying high-ranking government officials for our at-scale content reviewers, and to clarify when the publication of addresses and imagery of official residences of high-ranking government officials should be allowed. We expect to share a more robust update in 2023.

Oversight Board Recommendation: Meta should better explain, in the text of Facebook's Privacy Violations policy, when disclosing the city where a residence is located will suffice for the content to be removed, and when disclosing its neighborhood would be required for the same matter (e.g., by specifically referencing the population threshold at which sharing only the city as part of the content will no longer be considered violating). The Board will consider this implemented when Meta modifies its content policies.

(PAO on Sharing Residential Information #7)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	We are finalizing work to implement this recommendation, including modifying guidance and training materials to outline criteria for when disclosing the city or the neighborhood is sufficient for identifying where a residence is located. We hope to share an update in 2023.

Oversight Board Recommendation: Meta should explain, in the text of Facebook's Privacy Violations policy, its criteria for assessing whether the resident is sufficiently identified in the content. The Board will consider this implemented when Meta modifies its content policies.

(PAO on Sharing Residential Information #8)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In Progress
November 2022 Update	We are finalizing work to implement this recommendation, including modifying guidance and training materials to outline criteria for when a resident is sufficiently identified. We hope to share an update in 2023.

Oversight Board Recommendation: Meta should clarify the Hate Speech Community Standard and the guidance provided to reviewers, explaining that even implicit references to protected groups are prohibited by the policy when the reference would reasonably be understood. The Board will consider this recommendation implemented when Meta updates its Community Standards and Internal Implementation Standards to content reviewers to incorporate this revision.

(Video of an Edited Cartoon Depicting a Croatian City #1)

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
November 2022 Update	We are crafting language to add to our <u>Hate Speech</u> policy in the Community Standards that speaks to our approach to speech that is not explicit and requires additional context to interpret. We will report on our progress in a future Quarterly Update.

Oversight Board Recommendation: Meta should amend the Violent and Graphic Content Community Standard to allow videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses. This content should be allowed with a warning screen so that people are aware that content may be disturbing. The Board will consider this recommendation implemented when Meta updates the Community Standard.

(<u>Video Depicting a Civilian Victim of Violence in Sudan #1</u> (along with <u>Video Depicting a Civilian Victim of Violence in Sudan #2</u>)¹¹)

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

¹¹ The board issued a similar recommendation in recommendation #2 in the <u>Video Depicting a Civilian Victim of Violence in Sudan</u> case. We are tracking the progress of our work in response to this recommendation as part of our response to recommendation #1 in the Video Depicting a Civilian Victim of Violence in Sudan case.

November 2022 Update

We have started a policy development process to better understand how to strike a balance between documenting and raising awareness about human rights abuses against our responsibility to protect the dignity of the person depicted in the imagery and the potential harm caused by exposing users to this type of graphic content. As with past policy developments, we will be conducting research and seeking input from external stakeholders globally to inform potential options for this policy. We expect this policy development to continue into the next year, and hope to share details on our progress in a future Quarterly Update.

Oversight Board Recommendation: Meta should investigate why the December 2021 changes to the Dangerous Individuals and Organizations policy were not updated within the target time of six weeks, and ensure such delays or omissions are not repeated. The Board asks Meta to inform the Board within 60 days of the findings of its investigation, and the measures it has put in place to prevent translation delays in future.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #1)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	Complete
November 2022 Update	As shared in our <u>initial response</u> to this recommendation, we have investigated the cause of these translation delays and aligned on steps to prevent similar issues in the future. While the majority of our Community Standards policy language was translated within six weeks, the "Policy Rationale" section describing the aim of our Dangerous Individuals and Organizations policy was not translated concurrently. In order to prevent future translation delays, we will consolidate these drafts going forward. In addition, to help ensure that our translation updates are comprehensive and timely, we will complete an internal audit of our Community Standards translations on a quarterly basis. We will conduct our first translations audit in the coming months and now consider this recommendation complete.

Oversight Board Recommendation: Meta should make its public explanation of its two-track strikes system more comprehensive and accessible, especially for "severe strikes". It should include all policy violations that result in severe strikes, which account features can be limited as a result and specify applicable durations. Policies that result in severe strikes should also be clearly identified in the Community Standards, with a link to the "Restricting accounts" explanation of the strikes system. The Board asks Meta to inform the Board within 60 days of the updated Transparency Centre explanation of the strikes system, and the inclusion of the links to that explanation for all content policies that result in severe strikes.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #2)

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

November 2022 Update

As shared in our <u>initial response</u> to this recommendation, we agree that our public explanation of our penalty systems could be more comprehensive and accessible. We are currently reviewing our strikes system to make it more effective and proportionate, which includes assessing the types of policy violations that result in severe strikes, which account features can be limited as a result, and how long any restrictions will be in place. We will provide an update on this process in future Quarterly Updates.

Oversight Board Recommendation: Meta should narrow the definition of "praise" in the Known Questions guidance for reviewers, by removing the example of content that "seeks to make others think more positively about" a designated entity by attributing to them positive values or endorsing their actions. The Board asks Meta to provide the Board within 60 days with the full version of the updated Known Questions document for dangerous individuals and organizations.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #3)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	As shared in our <u>initial response</u> to this recommendation, we are currently conducting a policy development process to reevaluate our definition of "praise" in the Dangerous Individuals and Organizations policy to better promote user voice and safety on our platforms. This includes reassessing the definition of "praise" in the Community Standards and our internal Known Questions guidance. Because the next steps for this recommendation relate to our plans to address <u>Al Jazeera Post on Tensions Between Israel and Palestine Recommendation #1</u> , we will track future updates under that recommendation.

Oversight Board Recommendation: Meta should revise its internal Implementation Standards to make clear that the "reporting" allowance in the Dangerous Individuals Organizations policy allows for positive statements about designated entities as part of the reporting, and how to distinguish this from prohibited "praise." The Known Questions document should be expanded to make clear the importance of news reporting in situations of conflict or crisis and provide relevant examples, and that this may include positive statements about designated entities like the reporting on the Taliban in this case. The Board asks Meta to share the updated Implementation Standards with the Board within 60 days.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #4)

Previous Category	Implementing Fully
Updated Category	Implementing Fully
Current Status	In progress
November 2022 Update	In our <u>initial response</u> to this recommendation, we clarified that our Dangerous Individuals and Organizations policy currently allows content that includes references to designated dangerous organizations and individuals when that content is reporting on, condemning, or neutrally discussing them or their activities. However, there is also an ongoing policy

development process around how we define and enforce on "praise" under this policy. We will update our internal policy guidance to reflect any changes as a result of this policy development process and will update the board when this process is complete.

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.

(<u>Armenian People and the Armenian Genocide Recommendation #4</u>, (along with <u>PAO on Sharing Residential Information #10</u> and <u>PAO on Sharing Residential Information #14</u>)¹²)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	In our Q2 2022 Update, we noted that our product teams are working to identify key drivers of trust in appeals in order to improve appeals effectiveness and ensure that people on our platforms understand our content moderation decisions. In this effort, we are working on product developments that will allow appellants to indicate that they believe their content falls into a policy exception. To that end, our product and content designers have been working on a new design vision for enforcement and appeals experiences. These designs include tools that allow users to provide more context on why they are appealing an enforcement decision, and options that would enable users to indicate that their content should be allowed on the basis of a policy exception. These designs remain under development with a goal of inclusion in 2023 product roadmapping. As we've described previously, our work must account for the fact that not everyone is familiar with our policy exceptions, which could lead to inequitable enforcement outcomes. Additionally, we want to provide this new functionality without creating confusion or discouraging people from appealing. We will provide an update on the status of this recommendation in a future 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
November 2022 Update	In our Q2 2022 Quarterly Update, we articulated our ranking system for appeals and highlighted that we generally review appeals in the order we receive them with few

¹² The board issued similar recommendations in <u>PAO on Sharing Private Residential Information #10</u> and <u>PAO on Sharing Private Residential Information #14</u>. We are tracking the progress of our work in response to these recommendations as part of our response to recommendation #4 in the Armenian People and the Armenian Genocide case.

exceptions. We continue working to understand the tradeoffs of prioritizing certain appeals over others. In accordance with our response to <u>Armenian People and the Armenian Genocide Recommendation #4</u>, we are focused on refining the prerequisite work required to enable users to indicate that their content falls into a policy exception during the appeal. This will lead us towards solving more efficiently for how to prioritize. Our engineering, policy, and operations teams are still working to better understand these considerations and scope the subsequent product development. We will provide an update on the status of this ongoing work in a future Quarterly Update.

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
November 2022 Update	As previously explained in our Q2 Quarterly Update, we are currently undergoing a multi-stage process to understand and evaluate how we should prioritize appeals and how this impacts our legitimacy and fairness. We are currently working to evaluate user surveys related to this effort which will inform potential ranking models in the coming year. We will provide an update on the status of this recommendation in the next Quarterly Update.

Oversight Board Recommendation: Meta should conduct regular assessments on reviewer accuracy rates focused on the Restricted Goods and Services policy. The Board will consider this recommendation implemented when Meta shares the results of these assessments with the Board, including how these results will inform improvements to enforcement operations and policy development, and summarize the results in its quarterly Board transparency reports. Meta may consider if these assessments should be extended to reviewer accuracy rates under other Community

(Post Requesting Advice on Pharmaceutical Drugs #3)

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In Progress
November 2022 Update	As explained in our initial <u>response</u> to this recommendation, we currently collect and assess reviewer accuracy data on the basis of takedowns and restorations – including takedowns under our <u>Restricted Goods and Services policy</u> – to feed into continuous development of our classifiers and review protocols and policies. If we identify performance issues in the course of these continuous assessments, we attempt to identify the source of the issues and, depending on that information, involve the appropriate team to address it. In the interest of transparency around how we identify and address potential mistakes in the

enforcement of our Restricted Goods and Services Policy, we report on the amount of appealed content and content that is restored on Facebook and Instagram under that policy in our quarterly Community Standards Enforcement Report. As shared in our response to Punjabi Concern Over the RSS in India recommendation #3, our work defining reviewer accuracy metrics more broadly is underway. We hope to complete changes to our internal data logging infrastructure this year, but expect this to be a complex, long-term project. We will provide an update on the timeline for public reporting of these metrics in a future Quarterly Update.

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
November 2022 Update	As part of our broader privacy and strikes related work, we are continuing to assess criteria for content that is "clearly related to malicious action that created a risk of violence or harassment." We are continuing to evaluate ways to implement this recommendation that accounts for values such as safety and voice, and will keep the board updated on decisions and next steps in future Quarterly Updates.

Oversight Board Recommendation: Meta should assess the accuracy of reviewers enforcing the reporting allowance under the Dangerous Individuals and Organizations policy in order to identify systemic issues causing enforcement errors. The Board asks Meta to inform the Board within 60 days of the detailed results of its review of this assessment, or accuracy assessments Meta already conducts for its Dangerous Individuals and Organizations policy, including how the results will inform improvements to enforcement operations, including for HIPO.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #5)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In Progress
November 2022 Update	In our <u>initial response</u> to this recommendation, we stated that we plan to develop new tools that would allow us to gather more granular details about our enforcement of the DIO new reporting policy allowance. Based on proposed policy changes, we expect to explore updated decision tags that will allow human reviewers to indicate when content is non-violating because of this policy allowance. Additionally, we will consider looking into developing new classifiers that will help track this policy allowance enforcement going forward. We note that

this work is contingent on the outcome of an ongoing policy development process and we will update the board on our progress in future Quarterly Updates.

Oversight Board Recommendation: Meta should conduct a review of the HIPO ranker to examine if it can more effectively prioritize potential errors in the enforcement of allowances to the Dangerous Individuals and Organizations Policy. This should include examining whether the HIPO ranker needs to be more sensitive to news reporting content, where the likelihood of false-positive removals that impacts freedom of expression appears to be high. The Board asks Meta to inform the Board within 60 days of the results of its review and the improvements it will make to avoid errors of this kind in the future.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #6)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In progress
November 2022 Update	In our <u>initial response</u> to this recommendation, we share that we are reviewing the High Impact False Positive Override (HIPO) ranker for opportunities to improve the prioritization and effectiveness of the HIPO ranker to address false positive removals. We will update the board on the status of this work in future Quarterly Updates

Oversight Board Recommendation: Meta should enhance the capacity allocated to HIPO review across languages to ensure that more content decisions that may be enforcement errors receive additional human review. The Board asks Meta to inform the Board within 60 days of the planned capacity enhancements.

(Post from a News Outlet Discussing the Taliban Government in Afghanistan #7)

Previous Category	Assessing Feasibility
Updated Category	Assessing Feasibility
Current Status	In progress
November 2022 Update	In our <u>initial response</u> to this recommendation, we committed to exploring several improvements to increase HIPO review capacity. We are currently conducting internal experiments across multiple languages that appear to increase HIPO accuracy and expect that product improvements based on these findings will meaningfully increase human review capacity. We will share an update with the board on these improvements in future Quarterly Updates.

Oversight Board Recommendation: To improve Meta's ability to remove non-violating content from banks programmed to identify or automatically remove violating content, Meta should ensure that content with high rates of appeal and high rates of successful appeal is re-assessed for possible removal from its Media Matching Service banks. The Board will consider this recommendation implemented when Meta: (i) discloses to the Board the rates of appeal and successful appeal that trigger a review of Media Matching Service-banked content, and (ii) confirms publicly that these reassessment mechanisms are active for all its banks that target violating content.

(Cartoon Depicting Violence by the Police in Colombia #1)				
Previous Category	Implementing in Part			
Updated Category	Implementing in Part			
Current Status	In progress			
November 2022 Update	In our <u>initial response</u> to this recommendation, we shared that we will implement this recommendation with a gradual approach based on the complexity of the governance, enforcement, and maturity levels of individual banks. Some banking teams are already implementing this recommendation, while other banks are much newer and still in the process of training their auditing systems. Across all Media Matching Service (MMS) banks, however, we plan to implement several product and governance innovations to more effectively and efficiently remove incorrectly banked content. We will continue to report on our progress in the next Quarterly Update.			

Oversight Board Recommendation: To ensure that inaccurately banked content is quickly removed from Meta's Media Matching Service banks, Meta should set and adhere to standards that limit the time between when banked content is identified for re-review and when, if deemed non-violating, it is removed from the bank. The Board will consider this recommendation implemented when Meta: (i) sets and discloses to the Board its goal time between when a re-review is triggered and when the non-violating content is restored, and (ii) provides the Board with data demonstrating its progress in meeting this goal over the next year.

(Cartoon Depicting Violence by the Police in Colombia #2)

Previous Category	Implementing in Part
Updated Category	Implementing in Part
Current Status	In progress
November 2022 Update	As explained in our <u>initial response</u> to this recommendation, many of our individual MMS banking teams already have strict standards for the re-review and potential removal of flagged content within their banks. In response to this recommendation, we will work towards a more cohesive governance model for MMS banking to ensure that such standards exist for every bank. Because of widely differing use cases, policy types, and banking strategies, however, these standards will likely remain specific to each individual bank and will not be universal. We will also aim to share both existing and newly established time-to-review standards with the board. We will provide further updates on our progress in future Quarterly Updates.

IV. Index

Case	Recommendation	Updated Category	Status	Section	Page
Armenians in Azerbaijan	2020-003-FB-UA-1	Implementing in part	Complete	Transparency	18
Breast cancer symptoms and nudity	2020-004-IG-UA-2	Implementing fully	In progress	Policy	26
	2020-004-IG-UA-3	Implementing in part	Complete	Transparency Footnote 3	18
	2020-004-IG-UA-5	Implementing in part	In progress	Transparency	19
	2020-004-IG-UA-6	Assessing feasibility	In progress	Transparency	19
	2020-004-IG-UA-7	Assessing feasibility	In progress	Transparency	19
Nazi quote	2020-005-FB-UA-1	Implementing in part	Complete	Transparency Footnote 3	18
Former President Trump	2021-001-FB-FBR-18	Implementing in part	In progress	Transparency	20
Depiction of Zwarte Piet	2021-002-FB-UA-2	Implementing in part	Complete	<u>Transparency</u> Footnote 3	18
Punjabi Concerns Over the RSS in India	2021-003-FB-UA-3	Implementing in part	In progress	Transparency	21
January 2021	2021-004-FB-UA-3	Implementing in part	Complete	Policy	27
Protests in Russia	2021-004-FB-UA-4	Implementing in part	Complete	Policy Footnote 8	27
Armenian People	2021-005-FB-UA-4	Assessing feasibility	In progress	Enforcement	35
Genocide	2021-005-FB-UA-5	Assessing feasibility	In progress	Enforcement	35
Support of	2021-006-IG-UA-9	Implementing fully	In progress	Transparency	21
Abdullah Ocalan	2021-006-IG-UA-10	Implementing fully	In progress	Policy Footnote 7	26
	2021-006-IG-UA-11	Implementing fully	In progress	Transparency	22
Al Jazeera Post on Tensions Between	2021-009-FB-UA-1	Assessing feasibility	In progress	Policy	27
Israel and Palestine	2021-009-FB-UA-3	Implementing fully	Complete	Transparency	22

	2021-009-FB-UA-4	Implementing fully	In progress	Transparency Footnote 5	22
Post Depicting Protests in Colombia While Using a Slur	2021-010-FB-UA-4	Assessing feasibility	In progress	Transparency	23
South Africa Slur	2021-011-FB-UA-1	Implementing in part	Complete	Transparency Footnote 3	18
Depicting Indigenous Artwork and Discussing Residential Schools	2021-012-FB-UA-1	Implementing fully	In progress	<u>Transparency</u>	24
Post Discussing a Substance with	2021-013-IG-UA-1	Implementing fully	In progress	Policy Footnote 7	26
Psychoactive Properties	2021-013-IG-UA-2	Implementing in part	In progress	Transparency Footnote 3	18
	2021-013-IG-UA-3	Assessing feasibility	In progress	Policy	28
Post Requesting Advice on	2021-015-FB-UA-1	Implementing fully	Complete	Policy	28
Pharmaceutical Drugs	2021-015-FB-UA-2	Assessing feasibility	In progress	Enforcement	36
Diugs	2021-015-FB-UA-3	Implementing in part	In progress	Enforcement	36
Post Describing Sexual Violence	2021-016-FB-FBR-1	Implementing fully	In progress	Policy	29
Against Minors	2021-016-FB-FBR-2	Implementing fully	In progress	Policy	29
PAO on Sharing Private Residential	2021-001-FB-PAO-1	Implementing fully	In progress	Policy	29
Information	2021-001-FB-PAO-3	Implementing fully	In progress	Policy Footnote 9	29
	2021-001-FB-PAO-4	Implementing fully	In progress	Policy	30
	2021-001-FB-PAO-7	Assessing feasibility	In progress	Policy	30
	2021-001-FB-PAO-8	Implementing fully	In progress	Policy	31
	2021-001-FB-PAO-9	Implementing fully	In progress	Policy Footnote 7	26
	2021-001-FB-PAO-10	Assessing feasibility	In progress	Enforcement Footnote 11	35
	2021-001-FB-PAO-12	Assessing feasibility	In progress	Enforcement	37
	2021-001-FB-PAO-14	Assessing feasibility	In progress	Enforcement Footnote 11	35
	2021-001-FB-PAO-17	Implementing in part	Complete	Transparency Footnote 3	18
Video of an Edited	2022-001-FB-UA-1	Implementing in part	In progress	Policy	31

Cartoon Depicting a Croatian City	2022-001-FB-UA-2	Implementing fully	In progress	<u>Transparency</u>	24
Video Depicting a Civilian Victim of Violence in Sudan	2022-002-FB-MR-1	Assessing feasibility	In progress	Policy	31
	2022-002-FB-MR-2	Assessing feasibility	In progress	Policy Footnote 10	31
	2022-002-FB-MR-4	Assessing feasibility	In progress	Transparency Footnote 6	23
News Outlet	2022-005-FB-UA-1	Implementing in Full	Complete	Policy	32
Discussing the Taliban Government in Afghanistan	2022-005-FB-UA-2	Assessing Feasibility	In progress	Policy	32
III Algilaliistali	2022-005-FB-UA-3	Assessing Feasibility	In progress	Policy	33
	2022-005-FB-UA-4	Implementing in Full	In progress	Policy	33
	2022-005-FB-UA-5	Assessing Feasibility	In progress	Enforcement	37
	2022-005-FB-UA-6	Assessing Feasibility	In progress	Enforcement	38
	2022-005-FB-UA-7	Assessing Feasibility	In progress	Enforcement	38
Cartoon Depicting Violence by the Police in Colombia	2022-004-FB-UA-1	Implementing in Part	In progress	Enforcement	39
	2022-004-FB-UA-2	Implementing in Part	In progress	Enforcement	39
	2022-004-FB-UA-3	No Further Action	No further updates	Transparency	25