

On May 5, the Oversight Board upheld Facebook’s decision to suspend former President Donald Trump’s accounts, while also providing non-binding recommendations to improve our policies, how we enforce them, and our transparency reporting. We thank the Oversight Board for these recommendations. We identified 19 distinct recommendations and have responded to each of them below.

We carefully considered these recommendations and today are committing to fully implement 15 of them. We are also implementing one recommendation in part, still assessing two of them, and taking no further action on one of them. In determining how to respond to these recommendations, we consider the feasibility of implementation and the impact to our users. As with all of our most difficult decisions around content and policy, these choices require balancing difficult tradeoffs between issues that are often in conflict with one another. They also require understanding how our policies work not just on paper but also in practice. Where we cannot implement the board’s recommendations, we explain why it is not practical or feasible to do so.

This document serves as our response to each of the board’s recommendations.

We categorize our response to the board’s recommendations in the following areas:

- **Implementing fully:** Facebook agrees with the recommendation and has or will implement it in full.
- **Implementing in part:** Facebook agrees with the overall aim of the recommendation and has or will implement work related to the board's guidance.
- **Assessing feasibility:** Facebook is assessing the feasibility and impact of the recommendation and will provide further updates in the future.
- **No further action:** Facebook will not implement the recommendation, either due to a lack of feasibility or disagreement about how to reach the desired outcome.

## **1. Facebook should act quickly on posts made by influential users that pose a high probability of imminent harm.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Facebook often quickly reviews content posted by public figures that potentially violates our policies. We will continue to do so and find ways to improve this process while accounting for the complexity of analysis that is often required for this kind of content.

### **Considerations**

During especially high-risk events, such as elections and large-scale protests, Facebook regularly establishes an Integrity Product Operations Center (“IPOC”), which is a working group composed of subject matter experts from our product, policy, and operations teams. This structure allows these experts to more quickly surface, triage, investigate, and mitigate risks on the platform. When reviewing content from public figures, the team considers many factors, including how the content is being perceived by others, historical and cultural factors, and the consequences of various available enforcement actions. This review, which requires input from multiple teams, is described in greater detail in the response to the recommendation below.

For content that will likely be seen by many people, we may employ a cross check system to help ensure that we are applying our policies correctly, which we explain in our response to recommendation 10. We aim to strike a balance between the time it takes to examine the broader context discussed above and ensuring that violating content does not remain on our platform.

### **Next Steps**

We will work to shorten the deliberation process on these types of content decisions while maintaining appropriate rigor. As described in our response to recommendation 10, we will provide additional information about how we expedite our review of public figures in our Transparency Center.

## 2. Facebook should consider the context of posts by influential users when assessing a post's risk of harm.

### **Commitment Level**

Implementing fully

### **Our Commitment**

Facebook already considers the broader context of content from public figures in the course of our review, and we will continue to do so. Our consideration includes the relevant historical significance of statements, comments on the content that show how it is being understood, and how others are receiving similar content on our platform.

### **Considerations**

In some instances, when public figures are found to have posted potentially violating content, our teams evaluate it, including the perceived intent and potential impact. This assessment is often challenging. We strive to incorporate as much relevant context as possible into this analysis, including information such as comments, reactions, ongoing current events, relevant historical factors, and similar content posted by the user. Such contextual indicators are not necessarily available at every stage in the review process. For instance, content moderators working at scale have a more limited view of the content than specialized teams. We also may employ a cross check system for content that will likely be seen by many people as an additional safeguard to help ensure we are applying our policies correctly. We explain this system in our response to recommendation 10.

Our enforcement teams rely on the information available to them to make decisions quickly and at scale. We strive to balance (1) our ability to take action quickly, (2) protecting the privacy of our users, and (3) the surrounding context necessary to make enforcement decisions. Incorporating additional signals during our review could add helpful context providing that we can continue balancing speed, thoroughness, and the appropriate privacy safeguards.

### **Next Steps**

In some instances, Facebook considers the broader context of content from public figures in the course of our review. We will review our current processes to see how we

can best consider additional context when making enforcement decisions and what additional context may be helpful to assess risk at every stage in the review process.

---

### **3. Facebook should prioritize safety over expression when taking action on a threat of harm from influential users.**

#### **Commitment Level**

Implementing fully

#### **Our Commitment**

Facebook is, and has always been, committed to removing content where the risk of harm outweighs any public interest value. We will continue to prioritize public safety when making these judgements and will impose use restrictions and other feature blocks on accounts that violate our policies. We also quickly review content posted by public figures that potentially violates our policies so we can remove any violating content.

#### **Considerations**

Through our investments in AI, we have been able to find and remove more violating content proactively before people report it to us. To further reduce the risk of harm during high-risk events, we convene IPOCs (see our response to the first recommendation above for more information) to review and expedite content review during higher-risk events based on a variety of factors, including the severity of the potential violations and the reach of the content. When reviewing relevant content from public figures, we consider context such as the poster's intent, how the content is being received, and the consequences of various actions (see our response to the second recommendation).

In certain instances, policy determinations can take time to evaluate depending on the relevant policies and the public interest value in allowing content that may otherwise violate our Community Standards to remain on Facebook or Instagram. We explain our newsworthiness allowance in response to recommendation 11. These determinations generally require additional time to evaluate because of the complexity of the issues and the review required by multiple teams. And for content that will likely be seen by many

people and therefore potentially have a greater impact on public safety, we may employ a cross check system to help ensure we are applying our policies correctly, which we explain in our response to recommendation 10.

### **Next Steps**

We are working to enhance our automated tools to improve our proactive review of content that could potentially impact public safety.

---

- 4. Facebook should suspend the accounts of high government officials, such as heads of state, if their posts repeatedly pose a risk of harm.**

### **Our Commitment**

Implementing fully

### **Considerations**

See our response to recommendation 8.

### **Next Steps**

See our response to recommendation 8.

---

- 5. Facebook should suspend accounts of high government officials, such as heads of state, for a determinate period sufficient to protect against imminent harm. Periods of suspension should be long enough to deter misconduct and may, in appropriate cases, include account or page deletion.**

### **Our Commitment**

Implementing fully

### **Considerations**

See our response to recommendation 8.

### **Next Steps**

See our response to recommendation 8.

---

**6. Facebook should resist pressure from governments to silence their political opposition and consider the relevant political context, including off of Facebook and Instagram, when evaluating political speech from highly influential users.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Facebook already considers the broader context of content from public figures in the course of our review in some instances and undertakes accelerated review for public figures with adequate staff and resources. We will continue to do so. We are committed to exploring ways we can improve our external accountability as well as incorporate additional external feedback for our evaluation of political speech from public figures in accordance with our policies and especially during high risk events. In addition, we have a robust process for reviewing government reports alleging that content on Facebook violates local law.

### **Considerations**

We design our Community Standards to give everyone a voice while also keeping our community safe. We base our enforcement decisions on principled criteria. As described in our responses to the board's first, second, and third recommendations, we have processes in place to undertake accelerated review for public figures with a diverse global team from product, policy, and operations. We ensure that content reviewers are supported by teams with regional and linguistic expertise, including the context in which the speech is presented. And we will continue to provide adequate resources to support

that work. We also may employ a cross check system for content that will likely be seen by many people as an additional safeguard to help ensure we are applying our policies correctly. We explain this system in our response to recommendation 10.

When we receive a formal government report that content violates local law, we first review it against the Facebook Community Standards. If we determine that the content violates our policies, we remove it. If content does not violate our policies, in line with our commitments as a member of the [Global Network Initiative](#) and our [Corporate Human Rights Policy](#), we conduct a careful legal review to confirm whether the report is valid, as well as human rights due diligence.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions.

### **Next Steps**

We will continue to consider the broader context of content from public figures in the course of our review and undertake accelerated review for public figures with adequate staff and resources. We will look for additional ways to incorporate external feedback and hold ourselves more accountable for this review process.

7. Facebook should have a process that utilizes regional political and linguistic expertise, with adequate resourcing when evaluating political speech from highly influential users.

**Our Commitment**

Implementing fully

**Considerations**

See our response to recommendation 6.

**Next Steps**

See our response to recommendation 6.

---

8. Facebook should publicly explain the rules that it uses when it imposes account-level sanctions against influential users.

**Commitment Level**

Implementing fully

**Our Commitment**

Today, we are providing information in our Transparency Center about restricting accounts by public figures during civil unrest, which we created in response to the board's recommendations.

**Considerations**

When someone violates our Community Standards or Community Guidelines, we may impose a restriction on their account for a set period of time, or permanently disable, to eliminate the chance of additional violations and deter the user from committing future violations. We strive to keep restrictions proportionate to the violation the user committed.



Public figures often have broad influence across our platform and may therefore pose a greater risk of harm when they violate our Community Standards or Community Guidelines. Our standard restrictions of one to thirty days may not be proportionate to the violation or sufficient to reduce the risk of further harm in these cases, especially during ongoing violence or civil unrest.

Therefore, when determining the appropriate restriction for a public figure who has violated our Community Standards or Community Guidelines in ways that incite or celebrate ongoing violent disorder or civil unrest, we may consider: (1) the severity of the violation and the person's history on the platform, including current and past violations, (2) the public figure's potential influence over and relationship to the individuals engaged in violence, and (3) the severity of the violence and any related physical harm. During times of civil unrest and ongoing violence, we use these factors to determine the appropriate restriction length, ranging from one month to two years.

At the conclusion of the restriction period, we will assess whether the risk to public safety has receded. We will evaluate external factors, including instances of violence, restrictions on peaceful assembly, and other markers of global or civil unrest. If we determine that there is still a serious risk to public safety, we will extend the restriction for a set period of time and continue to re-evaluate until that risk has receded.

Once the public figure's restriction has expired and they regain access to the platform, to prevent repeated offenses they will be subject to heightened penalties. While most new violations will trigger a one month restriction from creating any content, more serious violations will merit a 2-year restriction from creating any content. In extreme cases we will permanently disable the account. If an account persistently posts any violating content, despite repeated warnings and restrictions, we will also disable the account.

### **Next Steps**

We have included information about [restricting accounts by public figures during civil unrest](#) in our Transparency Center.

**9. Facebook should assess the on-and-offline risk of harm before lifting an influential user’s account suspension.**

**Our Commitment**

Implementing fully

**Considerations**

See our response to recommendation 8.

**Next Steps**

See our response to recommendation 8.

---

**10. Facebook should document any exceptional processes that apply to influential users.**

**Commitment Level**

Implementing fully

**Our Commitment**

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

**Considerations**

At Facebook we moderate content that billions of people publish across our platform. We employ an additional review, called our cross check system, to help confirm we are applying our policies correctly for content that will likely be seen by many people. We have explained this process in [our Newsroom](#). We want to make clear that we remove content from Facebook, no matter who posts it, when it violates our Community Standards. There is only one exception — and that is for content that receives a newsworthiness allowance (see details in our response to the recommendation below).

Cross check simply means that we give some content from certain Pages or Profiles additional review. We often apply this process to ensure our policies are applied correctly for public figures and content on Facebook that will be seen by many people.

For additional information about our newsworthiness allowance and how we apply it to public figures, see our response to recommendation 11.

### **Next Steps**

We have documented our [cross check system](#) and [newsworthiness allowance](#) in our Transparency Center.

---

## **11. Facebook should more clearly explain its newsworthiness allowance.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Today, we are providing more information in our Transparency Center about our newsworthiness allowance and how we apply it. Next year we will also begin providing regular updates about when we apply our newsworthiness allowance. Finally, we are removing the presumption we announced in 2019 that speech from politicians is inherently of public interest.

### **Considerations**

We allow certain content that is newsworthy or important to the public interest to remain on our platform – even if it might otherwise violate our Community Standards. We may also limit other enforcement consequences, such as demotions, when it is in the public interest to do so. When making these determinations, however, we will remove content if the risk of harm outweighs the public interest.

We first described our newsworthiness allowance in a [Newsroom Post in 2016](#). In 2019 we provided [additional information](#) about this allowance and how and why we apply it to certain content. The [introduction to our Community Standards](#) has additional information

as well. We understand, however, that the board believes that there is still confusion about our newsworthiness allowance and so we are clarifying it today.

We grant our newsworthiness allowance to a small number of posts on our platform. We will begin publishing the rare instances when we apply it moving forward. Finally, when we assess content for newsworthiness, we will not treat content posted by politicians any differently from content posted by anyone else. Instead, we will simply apply our newsworthiness balancing test in the same way to all content, measuring whether the public interest value of the content outweighs the potential risk of harm by leaving it up.

### **Next Steps**

Today we posted additional information in our Transparency Center to clarify our [newsworthiness allowance](#) to address the board's concern that there is confusion about this allowance. Next year, we will also begin providing regular updates about the number of times we applied this allowance in our Community Standards Enforcement Reports. Finally we will no longer treat content from politicians as inherently of public interest.

### **Correction on the Application of the Newsworthiness Allowance to former President Trump**

We incorrectly told the board that we never issued a newsworthy allowance for any content on former President Trump's Facebook Page or Instagram Account. We discovered that we issued a newsworthy allowance for a video of a rally posted to his Page on August 15th, 2019 while compiling data on historical newsworthy allowances in response to the board's recommendations.

We applied the newsworthiness exception to an August 15th, 2019 [video on Mr. Trump's Page](#). During the video from a New Hampshire rally, Mr. Trump says: "That guy's got a serious weight problem. Go home. Start exercising."

This statement targets a private individual with an "attack through negative physical descriptions," which violates our Bullying & Harassment policy. Facebook decided to issue a newsworthy allowance because the violation was a small portion of a much longer video of a campaign rally that was being widely shared by media outlets for journalistic

purposes. At the time, we determined that there was a high public interest value for people to hear from an elected official running for re-election and a low risk of harm.

---

**12. In regard to cross check review for influential users, Facebook should clearly explain the rationale, standards, and processes of review, including the criteria to determine which pages and accounts are selected for inclusion.**

**Our Commitment**

Implementing fully

**Considerations**

See our response to recommendation 10.

**Next Steps**

See our response to recommendation 10.

---

**13. Facebook should report on the relative error rates and thematic consistency of determinations made through the cross check process compared with ordinary enforcement procedures.**

**Commitment Level**

No further action

**Our Commitment**

We will take no further action on this recommendation because it is not feasible to track this information.

### **Considerations**

While the board has requested details about the relative error rates of enforcement decisions made through cross check, we do not have systems in place to make this comparison. Our measurement accuracy systems are not designed to review the small number of decisions made through the cross check process.

### **Next Steps**

Because it is not feasible to track the requested information, we will take no further action on this recommendation.

---

## **14. Facebook should review its potential role in the election fraud narrative that sparked violence in the United States on January 6, 2021 and report on its findings.**

### **Commitment Level**

Implementing in part

### **Our Commitment**

We regularly review our policies and processes in response to real world events. We will continue to cooperate with law enforcement and any US government investigations related to the events on January 6. We have recently expanded our research initiatives to understand the effect that Facebook and Instagram have on elections, including by forming a partnership with nearly 20 outside academics to study this issue.

### **Considerations**

We are appalled by the events of January 6. We continually review whether and how we adjust our policies to combat misinformation and hate — and we agree it is appropriate this process takes into account the events of January 6. Our work to improve Facebook is never complete and we continually review our policies and practices in the face of evolving threats, changing tactics by malicious actors, and new situations in the world. Ultimately, though, we believe that independent researchers and our democratically elected officials are best positioned to complete an objective review of these events.

We have expanded research initiatives to understand the effect that Facebook and Instagram have on elections. Recently [we launched](#) a new research partnership with nearly 20 outside academics to look specifically at the role Facebook and Instagram played in the 2020 US election. This research will examine the impact of how people interact with our products, including content shared in News Feed and across Instagram, and the role of features like content ranking systems, with three guiding principles: independence, transparency, and consent. Regardless of what is discovered, Facebook will not restrict the researchers from publishing their findings. We also extended the data collection for this US 2020 partnership with independent academic researchers through the end of February 2021. This extension will allow researchers to better understand people's beliefs and opinions around events including the presidential transition, the violence at the Capitol on January 6, and the Inauguration.

Our Violence and Incitement policy prohibits content calling for or advocating violence, and we ban organizations and individuals that proclaim a violent mission under our Dangerous Organizations and Individuals policy. We believe our Dangerous Organizations and Individuals policy has long been the broadest and most aggressive in the industry, and we have used it to ban hate groups. Motivated by a range of indicators that suggested political violence in the United States was possible, in August 2020, we expanded this policy to address militarized social movements and violence-inducing conspiracy networks, such as QAnon. We have [provided information](#) about how we address movements and organizations tied to violence including updates about our takedown and enforcement efforts.

For example, from August to November 30, 2020, we removed about 3,200 Pages, 18,800 groups, 100 events, 23,300 Facebook profiles and 7,400 Instagram accounts for violating our policy against militarized social movements, most of them coming down prior to the election. At the same time, we also removed about 3,000 Pages, 9,800 groups, 420 events, 16,200 Facebook profiles, and 25,000 Instagram accounts for violating our policy against QAnon. Since then, we've continued to enforce this policy. As of January 12, 2021, we have identified over 890 militarized social movements to date and in total, removed about 3,400 Pages, 19,500 groups, 120 events, 25,300 Facebook profiles, and 7,500 Instagram accounts. We've also removed about 3,300 Pages, 10,500 groups, 510 events, 18,300 Facebook profiles, and 27,300 Instagram accounts for

violating our policy against QAnon. These groups are constantly working to avoid our enforcement and we will continue to study how they evolve in order to keep people safe.

The responsibility for January 6, 2021 lies with the insurrectionists and those who encouraged them, whose words and actions have no place on Facebook. We will continue to cooperate with law enforcement and any US government investigations related to the events on January 6. We also believe that an objective review of these events, including contributing societal and political factors, should be led by elected officials.

### **Next Steps**

We will continue to review how we can improve our policies and enforcement practices. We have expanded research initiatives to understand the effect that Facebook and Instagram have on elections, including by forming a partnership with nearly 20 outside academics to study this issue. We will continue to cooperate with law enforcement and US government investigations related to the events on January 6.

---

**15. Facebook should be clear in its Corporate Human Rights policy how it collects, preserves and shares information related to investigations and potential prosecutions, including how researchers can access that information.**

### **Commitment Level**

Assessing feasibility

### **Our Commitment**

We commit to reviewing our Corporate Human Rights Policy in response to this recommendation. We need time to evaluate the correct approach to data collection and preservation to facilitate lawful cooperation with diverse stakeholders given the complex legal and privacy issues in play. We will also explore how we can be more transparent about our protocols.



### **Considerations**

Facebook’s Human Rights Policy incorporates our data policy and law enforcement guidelines, which govern when law enforcement officials make related lawful requests. International privacy laws create a layer of complexity to collecting, preserving, and sharing user content and/or personal information or personally identifiable information. These laws also contain requirements about, among other things, data storage and data deletion requirements, which must be considered before we are able to fully address the board’s recommendation.

### **Next Steps**

We will review our Corporate Human Rights Policy to provide more clarity, and explore ways to improve protocols for information collection, preservation, and sharing. We will update our response to this recommendation with more information following that review.

---

**16. Facebook should explain in its Community Standards and Guidelines its strikes and penalties process for restricting profiles, pages, groups and accounts on Facebook and Instagram in a clear, comprehensive, and accessible manner.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Today we are publishing detailed information in our Transparency Center about our strikes and penalties. Our goal is to provide people with more information about our process for restricting profiles, pages, groups, and accounts on Facebook and Instagram.

### **Considerations**

Previously, we included information in our [Help Center](#) about what happens when Facebook removes people’s content. We recently added this information to our [Transparency Center](#) and today we are expanding it to include information about

how we impose strikes and how we calculate penalties so people can better understand our process. In providing this additional transparency, we want our global users to better understand the details of our strikes and penalties processes while avoiding including certain information that malicious actors could use to circumvent our enforcement systems.

### **Next Steps**

Today we [published](#) information about our strike system in our Transparency Center.

---

## **17. Facebook should tell users how many violations, strikes, and penalties they have, as well as the consequences of future violations.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Earlier this year, we launched “[Account Status](#)” on Facebook, an in-product experience to help every user understand the penalties Facebook applied to their accounts. It provides information about the penalties on a person’s account (currently active penalties as well as past penalties), including why we applied the penalty. In general, if people have a restriction on their account, they can see their history of certain violations, warnings, and restrictions their account might have, as well as how long this information will stay in Account Status on Facebook. We are committed to making further investments in this product to help people understand the details of our enforcement actions.

### **Considerations**

We have a range of penalties that we apply when users violate our Community Standards, and these can differ based on the severity or persistence of a violation. For more information about our strike and penalty process see our response to recommendation 16 above. In developing Account Status, we aim to provide as much information about what was removed, why, and the penalty applied without allowing malicious actors to circumvent our systems.

## **Next Steps**

We will continue working to be more transparent to our users to help them understand any penalties that are active on their account and why we have applied them. We published additional information about our strike system in our [Transparency Center today](#) in order to better explain it. We are also working to expand “Account Status” to Instagram. We will provide updates on our progress.

---

**18. 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.**

## **Commitment Level**

Assessing feasibility

## **Our Commitment**

We agree that sharing more information about enforcement actions would be beneficial and are assessing how best to do so in a way that is consistent and comprehensive.

## **Considerations**

There are several challenges to sharing data about enforcement actions broken down by region and country. First, in adversarial situations, country-level data is less reliable. For example, malicious actors who create fake accounts often mask the country where they are located, so it is difficult to report the accounts or content by where the person who posted content was located for data in categories such as fake accounts and spam. In addition, reporting by where the person who viewed the content is located is also challenging. When someone in one region posts content about another, which region takes priority? This challenge is particularly acute when it comes to groups and pages, where members, administrators, and subject matter often span countries.

### **Next Steps**

We are assessing how to report consistent and comprehensive data that provides meaningful transparency while also ensuring that the information is accurate. We will provide additional information after we complete our assessment.

---

**19. 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.**

### **Commitment Level**

Implementing fully

### **Our Commitment**

Facebook will develop a Crisis Policy Protocol which will be informed by various frameworks that we use to address risk, imminent harm, and integrity challenges. The protocol will focus on the threshold for when context specific policies are deployed, deactivated, and reassessed.

### **Considerations**

We have a series of policies and protocols designed to activate teams and centrally coordinate policy, operations, and product responses to integrity challenges stemming from real world events. These mechanisms include principled frameworks governing their use in situations where Facebook's scaled enforcement or business as usual processes do not fully address the risk posed by these integrity challenges. We assess and update these playbooks to incorporate lessons learned from responding to novel situations and prepare for future challenges.

### **Next Steps**

We plan to develop this protocol through the full Policy Forum process. We will provide additional information after we complete this process.