

FB-691QAMHJ

Case decision 2021-001-FB-FBR

Case summary

The Board has upheld Facebook's decision on January 7, 2021, to restrict then-President Donald Trump's access to posting content on his Facebook page and Instagram account.

However, it was not appropriate for Facebook to impose the indeterminate and standardless penalty of indefinite suspension. Facebook's normal penalties include removing the violating content, imposing a time-bound period of suspension, or permanently disabling the page and account.

The Board insists that Facebook review this matter to determine and justify a proportionate response that is consistent with the rules that are applied to other users of its platform. Facebook must complete its review of this matter within six months of the date of this decision. The Board also made policy recommendations for Facebook to implement in developing clear, necessary, and proportionate policies that promote public safety and respect freedom of expression.

About the case

Elections are a crucial part of democracy. On January 6, 2021, during the counting of the 2020 electoral votes, a mob forcibly entered the Capitol Building in Washington, D.C. This violence threatened the constitutional process. Five people died and many more were injured during the violence. During these events, then-President Donald Trump posted two pieces of content.

At 4:21 pm Eastern Standard Time, as the riot continued, Mr. Trump posted a video on

Facebook and Instagram:

I know your pain. I know you're hurt. We had an election that was stolen from us. It was a landslide election, and everyone knows it, especially the other side, but you have to go home now. We have to have peace. We have to have law and order. We have to respect our great people in law and order. We don't want anybody hurt. It's a very tough period of time. There's never been a time like this where such a thing happened, where they could take it away from all of us, from me, from you, from our country. This was a fraudulent election, but we can't play into the hands of these people. We have to have peace. So go home. We love you. You're very special. You've seen what happens. You see the way others are treated that are so bad and so evil. I know how you feel. But go home and go home in peace.

At 5:41 pm Eastern Standard Time, Facebook removed this post for violating its Community Standard on Dangerous Individuals and Organizations.

At 6:07 pm Eastern Standard Time, as police were securing the Capitol, Mr. Trump posted a written statement on Facebook:

These are the things and events that happen when a sacred landslide election victory is so unceremoniously viciously stripped away from great patriots who have been badly unfairly treated for so long. Go home with love in peace. Remember this day forever!

At 6:15 pm Eastern Standard Time, Facebook removed this post for violating its Community Standard on Dangerous Individuals and Organizations. It also blocked Mr. Trump from posting on Facebook or Instagram for 24 hours.

On January 7, after further reviewing Mr. Trump's posts, his recent communications off Facebook, and additional information about the severity of the violence at the Capitol, Facebook extended the block "indefinitely and for at least the next two weeks until the peaceful transition of power is complete."

On January 20, with the inauguration of President Joe Biden, Mr. Trump ceased to be the president of the United States.

On January 21, Facebook announced it had referred this case to the Board. Facebook

asked whether it correctly decided on January 7 to prohibit Mr. Trump's access to posting content on Facebook and Instagram for an indefinite amount of time. The

company also requested recommendations about suspensions when the user is a political leader.

In addition to the two posts on January 6, Facebook previously found five violations of its Community Standards in organic content posted on the Donald J. Trump Facebook page, three of which were within the last year. While the five violating posts were removed, no account-level sanctions were applied.

Key findings

The Board found that the two posts by Mr. Trump on January 6 severely violated Facebook's Community Standards and Instagram's Community Guidelines. "We love you. You're very special" in the first post and "great patriots" and "remember this day forever" in the second post violated Facebook's rules prohibiting praise or support of people engaged in violence.

The Board found that, in maintaining an unfounded narrative of electoral fraud and persistent calls to action, Mr. Trump created an environment where a serious risk of violence was possible. At the time of Mr. Trump's posts, there was a clear, immediate risk of harm and his words of support for those involved in the riots legitimized their violent actions. As president, Mr. Trump had a high level of influence. The reach of his posts was large, with 35 million followers on Facebook and 24 million on Instagram.

Given the seriousness of the violations and the ongoing risk of violence, Facebook was justified in suspending Mr. Trump's accounts on January 6 and extending that suspension on January 7.

However, it was not appropriate for Facebook to impose an 'indefinite' suspension.

It is not permissible for Facebook to keep a user off the platform for an undefined period, with no criteria for when or whether the account will be restored.

In applying this penalty, Facebook did not follow a clear, published procedure. 'Indefinite' suspensions are not described in the company's content policies.

Facebook's normal penalties include removing the violating content, imposing a time-bound period of suspension, or permanently disabling the page and account.

It is Facebook's role to create necessary and proportionate penalties that respond to severe violations of its content policies. The Board's role is to ensure that Facebook's rules and processes are consistent with its content policies, its values and its human rights commitments.

In applying a vague, standardless penalty and then referring this case to the Board to resolve, Facebook seeks to avoid its responsibilities. The Board declines Facebook's request and insists that Facebook apply and justify a defined penalty.

The Oversight Board's decision

The Oversight Board has upheld Facebook's decision to suspend Mr. Trump's access to post content on Facebook and Instagram on January 7, 2021. However, as Facebook suspended Mr. Trump's accounts 'indefinitely,' the company must reassess this penalty.

Within six months of this decision, Facebook must reexamine the arbitrary penalty it imposed on January 7 and decide the appropriate penalty. This penalty must be based on the gravity of the violation and the prospect of future harm. It must also be consistent with Facebook's rules for severe violations, which must, in turn, be clear, necessary and proportionate.

If Facebook decides to restore Mr. Trump's accounts, the company should apply its rules to that decision, including any changes made in response to the Board's policy recommendations below. In this scenario, Facebook must address any further violations promptly and in accordance with its established content policies.

A minority of the Board emphasized that Facebook should take steps to prevent the repetition of adverse human rights impacts and ensure that users who seek reinstatement after suspension recognize their wrongdoing and commit to observing the rules in the future.

When it referred this case to the Board, Facebook specifically requested "observations or recommendations from the Board about suspensions when the user is a political

leader.”

In a policy advisory statement, the Board made a number of recommendations to guide Facebook’s policies in regard to serious risks of harm posed by political leaders and other influential figures.

The Board stated that it is not always useful to draw a firm distinction between political leaders and other influential users, recognizing that other users with large audiences can also contribute to serious risks of harm.

While the same rules should apply to all users, context matters when assessing the probability and imminence of harm. When posts by influential users pose a high probability of imminent harm, Facebook should act quickly to enforce its rules. Although Facebook explained that it did not apply its ‘newsworthiness’ allowance in this case, the Board called on Facebook to address widespread confusion about how decisions relating to influential users are made. The Board stressed that considerations of newsworthiness should not take priority when urgent action is needed to prevent significant harm.

Facebook should publicly explain the rules that it uses when it imposes account-level sanctions against influential users. These rules should ensure that when Facebook imposes a time-limited suspension on the account of an influential user to reduce the risk of significant harm, it will assess whether the risk has receded before the suspension ends. If Facebook identifies that the user poses a serious risk of inciting imminent violence, discrimination or other lawless action at that time, another time-bound suspension should be imposed when such measures are necessary to protect public safety and proportionate to the risk.

The Board noted that heads of state and other high officials of government can have a greater power to cause harm than other people. If a head of state or high government official has repeatedly posted messages that pose a risk of harm under international human rights norms, Facebook should suspend the account for a period sufficient to protect against imminent harm. Suspension periods should be long enough to deter misconduct and may, in appropriate cases, include account or page deletion.

In other recommendations, the Board proposed that Facebook:

- Rapidly escalate content containing political speech from highly influential users to specialized staff who are familiar with the linguistic and political context. These staff should be insulated from political and economic interference, as well as undue influence.
- Dedicate adequate resourcing and expertise to assess risks of harm from influential accounts globally.
- Produce more information to help users understand and evaluate the process and criteria for applying the newsworthiness allowance, including how it applies to influential accounts. The company should also clearly explain the rationale, standards and processes of the cross check review, and report on the relative error rates of determinations made through cross check compared with ordinary enforcement procedures.
- Undertake a comprehensive review of Facebook's potential contribution to the narrative of electoral fraud and the exacerbated tensions that culminated in the violence in the United States on January 6. This should be an open reflection on the design and policy choices that Facebook has made that may allow its platform to be abused.
- Make clear in its corporate human rights policy how it collects, preserves and, where appropriate, shares information to assist in investigation and potential prosecution of grave violations of international criminal, human rights and humanitarian law.
- Explain its strikes and penalties process for restricting profiles, pages, groups and accounts in Facebook's Community Standards and Instagram's Community Guidelines.
- Include the number of profile, page, and account restrictions in its transparency reporting, with information broken down by region and country.
- Provide users with accessible information on how many violations, strikes and penalties have been assessed against them, and the consequences that will follow future violations.
- Develop and publish a policy that governs Facebook's response to crises or novel situations where its regular processes would not prevent or avoid imminent harm. This guidance should set appropriate parameters for such actions, including a requirement to review its decision within a fixed time.

*Case summaries provide an overview of the case and do not have precedential value.

Full case decision

In this case, Facebook asked the Board to answer two questions:

Considering Facebook’s values, specifically its commitment to voice and safety, did it correctly decide on January 7, 2021, to prohibit Donald J. Trump’s access to posting content on Facebook and Instagram for an indefinite amount of time?

In addition to the board’s determination on whether to uphold or overturn the indefinite suspension, Facebook welcomes observations or recommendations from the board about suspensions when the user is a political leader.

1. Decision summary

The Board upholds Facebook’s decision on January 7, 2021, to restrict then-President Donald Trump’s access to posting content on his Facebook page and Instagram account.

However, it was not appropriate for Facebook to impose the indeterminate and standardless penalty of indefinite suspension. Facebook’s normal penalties include removing the violating content, imposing a time-bound period of suspension, or permanently disabling the page and account.

The Board insists that Facebook review this matter to determine and justify a proportionate response that is consistent with the rules that are applied to other users of its platform. Facebook must complete its review of this matter within six months of the date of this decision. The Board also makes policy recommendations for Facebook to implement in developing clear, necessary, and proportionate policies that promote public safety and respect freedom of expression.

2. Case description

Elections are a crucial part of democracy. They allow people throughout the world to govern and to resolve social conflicts peacefully. In the United States of America, the Constitution says the president is selected by counting electoral college votes. On January 6, 2021, during the counting of the 2020 electoral votes, a mob forcibly

entered the Capitol where the electoral votes were being counted and threatened the constitutional process. Five people died and many more were injured during the violence.

Prior to January 6, then-President Donald Trump had asserted without evidence that the November 2020 presidential election had been stolen. Legal claims brought by Mr. Trump and others of election fraud were rejected in over 70 cases, and the then-Attorney General, after investigation, stated that there had been no fraud “on a scale that could have effected a different outcome in the election.” Nevertheless, Mr. Trump continued to make these unfounded claims, including through using Facebook, and referred to a rally planned for January 6:

1. On December 19, 2020, the Trump Facebook page posted: “Peter Navarro releases 36-page report alleging election fraud 'more than sufficient' to swing victory to Trump - A great report by Peter. Statistically impossible to have lost the 2020 Election. Big protest in D.C. on January 6th. Be there, will be wild!”
2. On January 1, 2021, the Trump Facebook page posted: “The BIG Protest Rally in Washington, D.C., will take place at 11.00 A.M. on January 6th. Locational details to follow. StopTheSteal!”

On the morning of January 6, 2021, Mr. Trump attended a rally near the White House and gave a speech. He continued to make unfounded claims that he won the election and suggested that Vice President Mike Pence should overturn President-elect Joe Biden’s victory, a power Mr. Pence did not have. He also stated, “we will stop the steal,” and “we’re going to the Capitol.”

Many of those attending the rally then marched to the U.S. Capitol Building, where they joined other protestors already gathered. Many of the protestors attacked Capitol security, violently entered the building, and rioted through the Capitol. Mr. Pence and other Members of Congress were placed at serious risk of targeted violence. Five people died and many were injured.

During these events, Mr. Trump posted a video and a statement to his Facebook page (which had at least 35 million followers), and the video was also shared to his Instagram account (which had at least 24 million followers). The posts stated the 2020 election was “stolen” and “stripped away.” The posts also praised and supported those who were at the time rioting inside the Capitol, while also calling on them to remain peaceful. Both the Facebook page and the Instagram account show a blue tick next to the page or account name, meaning that Facebook has confirmed that the account is

the page or account name, meaning that Facebook has confirmed that the account is the “authentic presence of the public figure” it represents.

In the one-minute video, posted at 4:21 pm Eastern Standard Time (EST), as the riot continued, Mr. Trump said:

I know your pain. I know you're hurt. We had an election that was stolen from us. It was a landslide election, and everyone knows it, especially the other side, but you have to go home now. We have to have peace. We have to have law and order. We have to respect our great people in law and order. We don't want anybody hurt. It's a very tough period of time. There's never been a time like this where such a thing happened, where they could take it away from all of us, from me, from you, from our country. This was a fraudulent election, but we can't play into the hands of these people. We have to have peace. So go home. We love you. You're very special. You've seen what happens. You see the way others are treated that are so bad and so evil. I know how you feel. But go home and go home in peace.

At 5:41 pm EST, Facebook removed this post for violating its Community Standard on Dangerous Individuals and Organizations.

Mr. Trump posted the following written statement at 6:07 pm EST, as police were securing the Capitol:

These are the things and events that happen when a sacred landslide election victory is so unceremoniously viciously stripped away from great patriots who have been badly unfairly treated for so long. Go home with love in peace. Remember this day forever!

At 6:15 pm EST, Facebook removed this post for violating its Community Standard on Dangerous Individuals and Organizations and imposed a 24-hour block on Mr. Trump's ability to post on Facebook or Instagram.

On January 7, 2021, after further reviewing Mr. Trump's posts, his recent communications off Facebook, and additional information about the severity of the violence at the Capitol, Facebook extended the block “indefinitely and for at least the next two weeks until the peaceful transition of power is complete.” Facebook cited Mr. Trump's “use of our platform to incite violent insurrection against a democratically elected government.”

In the days following January 6, some of the participants in the riot stated publicly that they did so at the behest of the president. One participant was quoted in the Washington Post (January 16, 2021): “I thought I was following my president. . . . He asked us to fly there. He asked us to be there. So I was doing what he asked us to do.” A video captured a rioter on the steps of the Capitol screaming at a police officer, “We were invited here! We were invited by the president of the United States!”

The District of Columbia declared a public emergency on January 6 and extended it until January 21 that same day. On January 27, the Department of Homeland Security (DHS) issued a National Terrorism Advisory System Bulletin warning of a “heightened threat environment across the United States, which DHS believes will persist in the weeks following the successful Presidential Inauguration.” It stated that “drivers to violence will remain through early 2021 and some [Domestic Violent Extremists] may be emboldened by the January 6, 2021 breach of the U.S. Capitol Building in Washington, D.C. to target elected officials and government facilities.”

While the posts that Facebook found to violate its content policies were removed, Mr. Trump’s Facebook page and Instagram account remain publicly accessible on Facebook and Instagram. There is no notice on the page or account of the restrictions that Facebook imposed. On January 21, 2021, Facebook announced that it had referred the case to the Oversight Board.

In addition to the two posts on January 6, 2021, Facebook previously found five violations of its Community Standards in organic content posted on the Donald J. Trump Facebook page, three within the last year. The five violating posts were removed, but no account-level sanctions were applied. In response to the Board’s question on whether any strikes had been applied, Facebook said that the page received one strike for a post in August 2020 which violated its COVID-19 Misinformation and Harm policy. Facebook did not explain why other violating content it had removed did not result in strikes.

Facebook has a “newsworthiness allowance” which allows content that violates its policies to remain on the platform, if Facebook considers the content “newsworthy and in the public interest.” Facebook asserted that it “has never applied the newsworthiness allowance to content posted by the Trump Facebook page or Instagram account.”

Responding to the Board’s questions, Facebook disclosed that “there were 20 pieces of content from Trump’s Facebook Page and Instagram Account that content reviewers or automation initially marked as violating Facebook’s Community Standards but were ultimately determined to not be violations.”

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

3. Authority and scope

The Oversight Board has the power to review a broad set of questions referred by Facebook (Charter Article 2, Section 1; Bylaws Article 2, Section 2.1). Decisions on these questions are binding and may include policy advisory statements with recommendations. These recommendations are non-binding but Facebook must respond to them (Charter Article 3, Section 4). The Board is an independent grievance mechanism to address disputes in a transparent and principled manner.

4. Relevant standards

Under the Oversight Board’s Charter, it must consider all cases in light of the following standards:

1. Facebook’s content policies:

Facebook has Community Standards that describe what users are not allowed to post on Facebook, and Instagram has Community Guidelines that describe what users are not allowed to post on Instagram.

Facebook’s Community Standard on Dangerous Individuals and Organizations prohibits “content that praises, supports, or represents events that Facebook designates as terrorist attacks, hate events, mass murders or attempted mass

murders, serial murders, hate crimes and violating events.” It also prohibits “content that praises any of the above organizations or individuals or any acts committed by them,” referring to hate organizations and criminal organizations, among others.

Instagram’s Community Guidelines state that “Instagram is not a place to support or praise terrorism, organized crime, or hate groups,” and provide a link to the Dangerous Individuals and Organizations Community Standard.

Facebook’s Community Standard on Violence and Incitement states it “remove[s] content, disable[s] accounts, and work[s] with law enforcement when [it] believe[s] there is a genuine risk of physical harm or direct threats to public safety.” The Standard specifically prohibits: “Statements advocating for high-severity violence” and “Any content containing statements of intent, calls for action, conditional or aspirational statements, or advocating for violence due to voting, voter registration or the administration or outcome of an election.” It also prohibits “Misinformation and unverifiable rumors that contribute to the risk of imminent violence or physical harm.”

Instagram’s Community Guidelines state that Facebook removes “content that contains credible threats” and “serious threats of harm to public and personal safety aren’t allowed.” Both sections include links to the Violence and Incitement Community Standard.

Facebook’s Terms of Service state that Facebook “may suspend or permanently disable access” to an account if it determines that a user has “clearly, seriously, or repeatedly” breached its terms or policies. The introduction to the Community Standards notes that “consequences for violating our Community Standards vary depending on the severity of the violation and the person's history on the platform.”

Instagram’s Terms of Use state that Facebook “can refuse to provide or stop providing all or part of the Service to you (including terminating or disabling your access to the Facebook Products and Facebook Company Products) immediately to protect our community or services, or if you create risk or legal exposure for us, violate these Terms of Use or our policies (including our Instagram Community Guidelines).” Instagram’s Community Guidelines state “Overstepping these boundaries may result in deleted content, disabled accounts, or other restrictions.”

II. Facebook’s values:

Facebook has five values outlined in the introduction to the Community Standards which it claims guide what is allowed on its platforms. Three of these values are “Voice,” “Safety,” and “Dignity.”

Facebook describes “Voice” as wanting “people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable. [...] Our commitment to expression is paramount, but we recognize that the Internet creates new and increased opportunities for abuse.”

Facebook describes “Safety” as Facebook’s commitment to “mak[e] Facebook a safe place” and states that “Expression that threatens people has the potential to intimidate, exclude or silence others and isn’t allowed on Facebook.”

Facebook describes “Dignity” as its belief that “all people are equal in dignity and rights” and states that it “expect[s] that people will respect the dignity of others and not harass or degrade others.”

III. Human rights standards:

On March 16, 2021, Facebook announced its corporate human rights policy, where it commemorated its commitment to respecting rights in accordance with the UN Guiding Principles on Business and Human Rights (UNGPs). The UNGPs, endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. As a global corporation committed to the UNGPs, Facebook must respect international human rights standards wherever it operates. The Oversight Board is called to evaluate Facebook’s decision in view of international human rights standards as applicable to Facebook.

The Board analyzed Facebook’s human rights responsibilities in this case by considering human rights standards including:

- The right to freedom of expression: International Covenant on Civil and Political Rights (ICCPR), Articles 19 and 20; as interpreted in General Comment No. 34, Human Rights Committee (2011) (General Comment 34); the Rabat Plan of Action, OHCHR, (2012); UN Special Rapporteur on freedom of opinion and expression report A/HRC/38/35 (2018); Joint Statement of international freedom of expression monitors on COVID-19 (March, 2020).
- The right to life: ICCPR Article 6.

- The right to security of person, ICCPR Article 9, para. 1

- The right to security of person: ICCPR Article 9, para. 1.
- The right to non-discrimination: ICCPR Articles 2 and 26; International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Articles 1 and 4.
- Participation in public affairs and the right to vote: ICCPR Article 25.
- The right to remedy: ICCPR Article 2; General Comment No. 31, Human Rights Committee (2004) (General Comment 31); UNGPs, Principle 22.

5. Content creator's statement

When Facebook refers a case to the Board, the Board gives the person responsible for the content the opportunity to submit a statement. In this case, a statement to the Board was submitted on Mr. Trump's behalf through the American Center for Law and Justice and a page administrator. This statement requests that the Board "reverse Facebook's indefinite suspension of the Facebook account of former U.S. President Donald Trump."

The statement discusses the posts removed from Facebook and Instagram on January 6, 2021, as well as Mr. Trump's speech earlier that day. It states that the posts "called for those present at and around the Capitol that day to be peaceful and law abiding, and to respect the police" and that it is "inconceivable that either of those two posts can be viewed as a threat to public safety, or an incitement to violence." It also states that "It is stunningly clear that in his speech there was no call to insurrection, no incitement to violence, and no threat to public safety in any manner," and describes a "total absence of any serious linkage between the Trump speech and the Capitol building incursion."

The statement also discusses Facebook's reasons for imposing the restrictions. It states that as "nothing Mr. Trump said to the rally attendees could reasonably be interpreted as a threat to public safety," Facebook's basis for imposing restrictions cannot be safety-related. It also states that "any content suspected of impacting safety must have a direct and obvious link to actual risk of violence." The statement further describes that the terms "fight" or "fighting" used during the rally speech "were linked to a call for lawful political and civic engagement," and concludes "those words were neither intended, nor would be believed by any reasonable observer or listener to be a call for violent insurrection or lawlessness."

The statement also addresses the "Capitol incursion." It states that "all genuine Trump

political supporters were law-abiding" and that the incursion was "certainly influenced, and most probably ignited by outside forces." It describes a federal complaint against members of the Oath Keepers, and states the group was "in no way associated with Mr. Trump or his political organization." It then states the Oath Keepers were "parasitically using the Trump rally and co-opting the issue of the Electoral College debate for their own purposes."

The statement argues that Mr. Trump's rally speech did not violate the Dangerous Organizations and Individuals Community Standard because "none of those categories fit this case" and "Mr. Trump's political speech on January 6th never 'proclaim[ed] a violent mission,' a risk that lies at the very center of the Facebook policy." It also states the Violence and Incitement Community Standard "fail[s] to support the suspension of the Trump Facebook account" because the two posts "merely called for peace and safety" and "none of the words in Mr. Trump's speech, when considered in their true context, could reasonably be construed as incitement to violence or lawlessness." It also cites Facebook's referral to the Board mentioning the "peaceful transfer of power" and states this "new ad hoc rule on insuring [sic] peaceful governmental transitions is not just overly vague, it was non-existent until after the events that Facebook used to justify it."

The statement also argues that the Board should "defer to American law in this appeal" and discusses the international law standards for restricting the right to freedom of expression, of legality, legitimate aim, and necessity and proportionality, with each element interpreted by reference to United States constitutional law. On legality, the statement cites protection of hyperbole and false statements of fact and Facebook's importance to public discourse. It states that "employing content decisions based on what seems 'reasonable,' or how a 'reasonable person' would react to that content is not enough" and Facebook should "consider a much higher bar." It states that the Supreme Court requires strict scrutiny for laws that burden political speech and that Facebook has market dominance. It also discusses constitutional standards for incitement to violence. On legitimate aim, it states that preserving public safety is a legitimate aim, but Mr. Trump's speech did not present safety concerns. On necessity and proportionality, it denies the validity of the restrictions and states the penalty was disproportionate.

The statement concludes with suggestions for the Board's policy recommendations on suspensions when the user is a political leader. It argues that the Board should "defer

to the legal principles of the nation state in which the leader is, or was governing.” It then described multiple exceptions to this deference based on assessments of rule of law, guarantees of rights, processes for law making, processes of judicial review, and the existence of relevant legal principles in particular countries.

6. Facebook’s explanation of its decision

For each case, Facebook provides an explanation of its actions to the Board, and the Board asks Facebook questions to clarify further information it requires to make its decision. In this case, Facebook states that it removed the two pieces of content posted on January 6, 2021, for violating the Dangerous Individuals and Organizations Community Standard. Specifically, the content was removed for violating “its policy prohibiting praise, support, and representation of designated Violent Events.” Facebook also stated it contained “a violation of its Dangerous Individuals and Organizations policy prohibiting praise of individuals who have engaged in acts of organized violence.” The company notes that its Community Standards clearly prohibit “content that expresses support or praise for groups, leaders or individuals involved in” activities such as terrorism, organized violence or criminal activity, and that this includes organized assault as well as planned acts of violence attempting to cause injury to a person with the intent to intimidate a government in order to achieve a political aim.

Facebook notes that its assessment reflected both the letter of its policy and the surrounding context in which the statement was made, including the ongoing violence at the Capitol. It says that while Mr. Trump did ask people in his video to “go home in peace,” he also reiterated allegations that the election was fraudulent and suggested a common purpose in saying, “I know how you feel.” Given the ongoing instability at the time of his comments and the overall tenor of his words, Facebook concludes that “We love you. You’re very special” was intended as praise of people who were breaking the law by storming the Capitol. It also believes the second post to contain praise of the event, as Mr. Trump referred to those who stormed the Capitol as “great patriots,” and urged people to “[r]emember this day forever.”

Facebook notes that it regularly limits the functionality of Facebook pages and profiles and Instagram accounts which repeatedly or severely violate its policies. Where it concludes that there is an “urgent and serious safety risk,” Facebook “goes beyond its

standard enforcement protocols to take stronger actions against users and pages engaged in violating behavior.” In such cases, Facebook states that its enforcement actions remain grounded in its Community Standards and Instagram’s Community Guidelines. It states that it “evaluates all available enforcement tools, including permanent bans, before deciding which is the most appropriate to employ in the unique circumstance. In cases where Facebook must make an emergency decision that has widespread interest, it endeavors to share its decision and its reasoning with the public, often through a post in its Newsroom.”

Facebook states that it usually does not block the ability of pages to post or interact with content, but removes pages which severely or repeatedly violate Facebook’s policies. However, Facebook notes that its enforcement protocols for profiles, including feature blocks, may also be applied to Facebook pages when they are used in a person’s singular voice, as with the Donald J. Trump page. In this case, Facebook states that, in line with its standard enforcement protocols, it initially imposed a 24-hour block on the ability to post from the Facebook page and Instagram account. After further assessing the evolving situation and emerging details of the violence at the Capitol, Facebook concluded that the 24-hour ban was not sufficient to address “the risk that Trump would use his Facebook and Instagram presence to contribute to a risk of further violence.”

Facebook notes that it maintained the indefinite suspension after Mr. Biden’s inauguration partly due to analysis that violence connected to Mr. Trump had not passed. It cites National Terrorism Advisory System Bulletin issued on January 27 by the Department of Homeland Security (DHS) that described a “heightened threat environment across the United States, which DHS believes will persist in the weeks following the successful Presidential Inauguration” and that “drivers to violence will remain through early 2021 and some [Domestic Violent Extremists] may be emboldened by the January 6, 2021, breach of the U.S. Capitol Building in Washington, D.C. to target elected officials and government facilities.” Facebook notes that even when the risk of violence has diminished, it may be appropriate to permanently block Mr. Trump’s ability to post based on the seriousness of his violations on January 6, his continued insistence that Mr. Biden’s election was fraudulent, his sharing of other misinformation, and the fact that he is no longer president.

Facebook states that its decision was “informed by Article 19 of the ICCPR, and U.N. General Comment No. 34 on freedom of expression, which permits necessary and

proportionate restrictions of freedom of expression in situations of public emergency that threatens the life of the nation. In this case, the District of Columbia was operating under a state of emergency that had been declared to protect the U.S. Capitol

complex.” Facebook notes that it also took into account the six contextual factors from the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred. The Rabat Plan of Action was developed by experts with the support of the United Nations to guide states in addressing when advocacy of racial, religious or national hatred that incites discrimination, hostility or violence is so serious that resort to state-imposed criminal sanctions is appropriate, while protecting freedom of expression, in line with states’ obligations under Article 19 and Article 20, para. 2 of the ICCPR.

Facebook argues that the events of January 6 represented an unprecedented threat to the democratic processes and constitutional system of the United States. While Facebook asserts that it strives to act proportionately and accountably in curtailing public speech, given the unprecedented and volatile circumstances, Facebook believes it should retain operational flexibility to take further action including a permanent ban.

In this case, the Board asked Facebook 46 questions, and Facebook declined to answer seven entirely, and two partially. The questions that Facebook did not answer included questions about how Facebook’s news feed and other features impacted the visibility of Mr. Trump’s content; whether Facebook has researched, or plans to research, those design decisions in relation to the events of January 6, 2021; and information about violating content from followers of Mr. Trump’s accounts. The Board also asked questions related to the suspension of other political figures and removal of other content; whether Facebook had been contacted by political officeholders or their staff about the suspension of Mr. Trump’s accounts; and whether account suspension or deletion impacts the ability of advertisers to target the accounts of followers. Facebook stated that this information was not reasonably required for decision-making in accordance with the intent of the Charter; was not technically feasible to provide; was covered by attorney/client privilege; and/or could not or should not be provided because of legal, privacy, safety, or data protection concerns.

7. Third-party submissions

The Oversight Board received 9,666 public comments related to this case. Eighty of

the comments were submitted from Asia Pacific and Oceania, seven from Central and South Asia, 136 from Europe, 23 from Latin America and the Caribbean, 13 from the Middle East and North Africa, 19 from Sub-Saharan Africa, and 9,388 from the United States and Canada.

The submissions cover the following themes, which include issues that the Board specifically asked about in its call for public comments:

- Facebook’s indefinite suspension of Mr. Trump’s accounts, its possible compliance with the company’s responsibilities to respect freedom of expression and human rights, and if alternative measures should have been taken.
- Facebook’s policies and practices on assessing off-Facebook context in enforcing its Community Standards, particularly if that content may incite violence.
- Issues involving the clarity of Facebook’s rules for disabling accounts.
- Facebook’s global content policies with respect to political candidates, office holders, and former office holders, including the relevance of Facebook’s “newsworthiness” allowance and the public’s right to information.
- Concerns regarding the consistency in the enforcement of Facebook’s Community Standards based on political bias.
- Concerns about the enforcement of Facebook’s Community Standards with respect to Mr. Trump’s previous posts, including those that may have contributed to harm towards certain groups of people and the spread of misinformation.
- Whether or not Mr. Trump’s expression, both prior to and on January 6, constituted incitement to violence.
- The outcome of the U.S. election and the Trump presidency.

To read public comments submitted for this case, please click [here](#).

8. Oversight Board analysis

8.1 Compliance with content policies

The Board agrees with Facebook’s decision that the two posts by Mr. Trump on January 6 violated Facebook’s Community Standards and Instagram’s Community Guidelines. Facebook’s Community Standard on Dangerous Individuals and Organizations says that users should not post content “expressing support or praise for group leaders or individuals involved in” violating events. Facebook designated

for groups, leaders, or individuals involved in violating events. Facebook designated the storming of the Capitol as a “violating event” and noted that it interprets violating events to include designated “violent” events.

At the time the posts were made, the violence at the Capitol was underway. Both posts praised or supported people who were engaged in violence. The words “We love you. You’re very special” in the first post and “great patriots” and “remember this day forever” in the second post amounted to praise or support of the individuals involved in the violence and the events at the Capitol that day.

The Board notes that other Community Standards may have been violated in this case, including the Standard on Violence and Incitement. Because Facebook’s decision was not based on this Standard and an additional finding of violation would not affect the outcome of this proceeding, a majority of the Board refrains from reaching any judgment on this alternative ground. The decision upholding Facebook’s imposition of restrictions on Mr. Trump’s accounts is based on the violation of the Dangerous Individuals and Organizations Community Standard.

A minority of the Board would consider the additional ground and find that the Violence and Incitement Standard was violated. The minority would hold that, read in context, the posts stating the election was being “stolen from us” and “so unceremoniously viciously stripped,” coupled with praise of the rioters, qualifies as “calls for actions,” “advocating for violence” and “misinformation and unverifiable rumors that contribute[d] to the risk of imminent violence or physical harm” prohibited by the Violence and Incitement Community Standard.

The Board finds that the two posts severely violated Facebook policies and concludes that Facebook was justified in restricting the account and page on January 6 and 7. The user praised and supported people involved in a continuing riot where people died, lawmakers were put at serious risk of harm, and a key democratic process was disrupted. Moreover, at the time when these restrictions were extended on January 7, the situation was fluid and serious safety concerns remained. Given the circumstances, restricting Mr. Trump’s access to Facebook and Instagram past January 6 and 7 struck an appropriate balance in light of the continuing risk of violence and disruption. As discussed more fully below, however, Facebook’s decision to make those restrictions “indefinite” finds no support in the Community Standards and violates principles of freedom of expression.

The Board notes that there is limited detailed public information on the cross check system and newsworthiness allowance. Although Facebook states the same rules apply to high-profile accounts and regular accounts, different processes may lead to different substantive outcomes. Facebook told the Board that it did not apply the newsworthiness allowance to the posts at issue in this case. Unfortunately, the lack of transparency regarding these decision-making processes appears to contribute to perceptions that the company may be unduly influenced by political or commercial considerations.

8.2 Compliance with Facebook's values

The analysis above is consistent with Facebook's stated values of "Voice" and "Safety." For the reasons stated in this opinion, in this case the protection of public order justified limiting freedom of expression.

A minority believes it is particularly important to emphasize that "Dignity" was also relevant. Facebook relates "Dignity" to equality and that people should not "harass or degrade" others. The minority considers below that previous posts on the platform by Mr. Trump contributed to racial tension and exclusion and that this context was key to understanding the impact of Mr. Trump's content. Having dealt with this case on other grounds, the majority does not comment on these posts.

8.3 Compliance with Facebook's human rights responsibilities

The Board's decisions do not concern the human rights obligations of states or application of national laws, but focus on Facebook's content policies, its values and its human rights responsibilities as a business. The UN Guiding Principles on Business and Human Rights, which Facebook has endorsed (See Section 4), establish what businesses should do on a voluntary basis to meet these responsibilities. This includes avoiding causing or contributing to human rights harms, in part through identifying possible and actual harms and working to prevent or address them (UNGP Principles 11, 13, 15, 18). These responsibilities extend to harms caused by third parties (UNGP Principle 19).

Facebook has become a virtually indispensable medium for political discourse, and especially so in election periods. It has a responsibility both to allow political expression and to avoid serious risks to other human rights. Facebook, like other digital

platforms and media companies, has been heavily criticized for distributing misinformation and amplifying controversial and inflammatory material. Facebook's

human rights responsibilities must be understood in the light of those sometimes competing considerations.

The Board analyzes Facebook's human rights responsibilities through international standards on freedom of expression and the rights to life, security, and political participation. Article 19 of the ICCPR sets out the right to freedom of expression. Article 19 states that "everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." The Board does not apply the First Amendment of the U.S. Constitution, which does not govern the conduct of private companies. However, the Board notes that in many relevant respects the principles of freedom of expression reflected in the First Amendment are similar or analogous to the principles of freedom of expression in ICCPR Article 19.

Political speech receives high protection under human rights law because of its importance to democratic debate. The UN Human Rights Committee provided authoritative guidance on Article 19 ICCPR in General Comment No. 34, in which it states that "free communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential" (para. 20).

Facebook's decision to suspend Mr. Trump's Facebook page and Instagram account has freedom of expression implications not only for Mr. Trump but also for the rights of people to hear from political leaders, whether they support them or not. Although political figures do not have a greater right to freedom of expression than other people, restricting their speech can harm the rights of other people to be informed and participate in political affairs. However, international human rights standards expect state actors to condemn violence (Rabat Plan of Action), and to provide accurate information to the public on matters of public interest, while also correcting misinformation (2020 Joint Statement of international freedom of expression monitors on COVID-19).

International law allows for expression to be limited when certain conditions are met.

Any restrictions must meet three requirements – rules must be clear and accessible, they must be designed for a legitimate aim, and they must be necessary and proportionate to the risk of harm. The Board uses this three-part test to analyze

Facebook's actions when it restricts content or accounts. First Amendment principles under U.S. law also insist that restrictions on freedom of speech imposed through state action may not be vague, must be for important governmental reasons and must be narrowly tailored to the risk of harm.

I. Legality (clarity and accessibility of the rules)

In international law on freedom of expression, the principle of legality requires that any rule used to limit expression is clear and accessible. People must be able to understand what is allowed and what is not allowed. Equally important, rules must be sufficiently clear to provide guidance to those who make decisions on limiting expression, so that these rules do not confer unfettered discretion, which can result in selective application of the rules. In this case, these rules are Facebook's Community Standards and Instagram's Community Guidelines. These policies aim to set out what people cannot post, and Facebook's policies on when it can restrict access to Facebook and Instagram accounts.

The clarity of the Standard against praise and support of Dangerous Individuals and Organizations leaves much to be desired, as the Board noted in a prior decision (case [2020-005-FB-UA](#)). The UN Special Rapporteur on Freedom of Expression has also raised concerns about the vagueness of the Dangerous Individuals and Organizations Standard (A/HRC/38/35, para 26, footnote 67). As the Board has noted previously in case [2020-003-FB-UA](#), there may be times in which certain wording may raise legality concerns, but as applied to a particular case those concerns are not warranted. Any vagueness under the terms of the Standard does not render its application to the circumstances of this case doubtful. The January 6 riot at the Capitol fell squarely within the types of harmful events set out in Facebook's policy, and Mr. Trump's posts praised and supported those involved at the very time the violence was going on, and while Members of Congress were calling on him for help. In relation to these facts, Facebook's policies gave adequate notice to the user and guidance to those enforcing the rule.

With regard to penalties for violations, the Community Standards and related information about account restrictions are published in various sources, including the

Terms of Service, the introduction to the Community Standards, the Community Standard on Account Integrity and Authentic Identity, the [Facebook Newsroom](#), and the Facebook Help Center. As noted in case [2020-006-FB-FBR](#) the Board reiterates that the patchwork of applicable rules makes it difficult for users to understand why and when Facebook restricts accounts, and raises legality concerns.

While the Board is satisfied that the Dangerous Individuals and Organizations Standard is sufficiently clear under the circumstances of this case to satisfy clarity and vagueness norms of freedom of speech, Facebook's imposition of an "indefinite" restriction is vague and uncertain. "Indefinite" restrictions are not described in the Community Standards and it is unclear what standards would trigger this penalty or what standards will be employed to maintain or remove it. Facebook provided no information of any prior imposition of indefinite suspensions in any other cases. The Board recognizes the necessity of some discretion on Facebook's part to suspend accounts in urgent situations like that of January, but users cannot be left in a state of uncertainty for an indefinite time.

The Board rejects Facebook's request for it to endorse indefinite restrictions, imposed and lifted without clear criteria. Appropriate limits on discretionary powers are crucial to distinguish the legitimate use of discretion from possible scenarios around the world in which Facebook may unduly silence speech not linked to harm or delay action critical to protecting people.

II. Legitimate aim

The requirement of legitimate aim means that any measure restricting expression must be for a purpose listed in Article 19, para. 3 of the ICCPR, and this list of aims is exhaustive. Legitimate aims include the protection of public order, as well as respect for the rights of others, including the rights to life, security, and to participate in elections and to have the outcome respected and implemented. An aim would not be legitimate where used as a pretext for suppressing expression, for example, to cite the aims of protecting security or the rights of others to censor speech simply because it is disagreeable or offensive (General Comment No. 34, paras. 11, 30, 46, 48). Facebook's policy on praising and supporting individuals involved in "violating events," violence or criminal activity was in accordance with the aims above.

III. Necessity and proportionality

The requirement of necessity and proportionality means that any restriction on expression must, among other things, be the least intrusive way to achieve a legitimate aim (General Comment No. 34, para. 34).

The Board believes that, where possible, Facebook should use less restrictive measures to address potentially harmful speech and protect the rights of others before resorting to content removal and account restriction. At a minimum, this would mean developing effective mechanisms to avoid amplifying speech that poses risks of imminent violence, discrimination, or other lawless action, where possible and proportionate, rather than banning the speech outright.

Facebook stated to the Board that it considered Mr. Trump's "repeated use of Facebook and other platforms to undermine confidence in the integrity of the election (necessitating repeated application by Facebook of authoritative labels correcting the misinformation) represented an extraordinary abuse of the platform." The Board sought clarification from Facebook about the extent to which the platform's design decisions, including algorithms, policies, procedures and technical features, amplified Mr. Trump's posts after the election and whether Facebook had conducted any internal analysis of whether such design decisions may have contributed to the events of January 6. Facebook declined to answer these questions. This makes it difficult for the Board to assess whether less severe measures, taken earlier, may have been sufficient to protect the rights of others.

The crucial question is whether Facebook's decision to restrict access to Mr. Trump's accounts on January 6 and 7 was necessary and proportionate to protect the rights of others. To understand the risk posed by the January 6 posts, the Board assessed Mr. Trump's Facebook and Instagram posts and off-platform comments since the November election. In maintaining an unfounded narrative of electoral fraud and persistent calls to action, Mr. Trump created an environment where a serious risk of violence was possible. On January 6, Mr. Trump's words of support to those involved in the riot legitimized their violent actions. Although the messages included a seemingly perfunctory call for people to act peacefully, this was insufficient to defuse the tensions and remove the risk of harm that his supporting statements contributed to. It was appropriate for Facebook to interpret Mr. Trump's posts on January 6 in the context of escalating tensions in the United States and Mr. Trump's statements in other media and at public events.

As part of its analysis, the Board drew upon the six factors from the Rabat Plan of Action to assess the capacity of speech to create a serious risk of inciting discrimination, violence, or other lawless action:

- **Context:** The posts were made during a time of high political tension centered on the unfounded claim that the November 2020 presidential election had been stolen. The Trump campaign had raised these claims in court, but with little or no evidence, and they were consistently rejected. Mr. Trump nonetheless continued to assert these claims on social media, including Facebook and Instagram, using his authoritative status as head of state to lend them credibility. He encouraged supporters to come to the nation's capital on January 6 to "StoptheSteal," suggesting that the events would be "wild." On January 6, Mr. Trump urged supporters to march to the Capitol building to challenge the counting of the electoral votes. At the time of the posts, severe violence was continuing. When the restrictions were extended on January 7, the situation remained volatile. Among other indicators of the context, the District of Columbia took steps to warn of a heightened risk of violence surrounding the events at the Capitol.
- **Status of the speaker:** Mr. Trump's identity as president of the United States and a political leader gave his Facebook and Instagram posts a high degree of influence. The Board notes that as president, Mr. Trump had credibility and authority with members of the public, which contributed to the events of January 6. Mr. Trump's status as head of state with a high position of trust not only imbued his words with greater force and credibility but also created risks that his followers would understand they could act with impunity.
- **Intent:** The Board is not in a position to conclusively assess Mr. Trump's intentions. The possibility of violence linked to Mr. Trump's statements was clear, and the Board considered that he likely knew or should have known that these communications would pose a risk of legitimizing or encouraging violence.
- **Content and form:** The two posts on January 6 praised and supported rioters, even though they called on them to go home peacefully. The posts also reiterated the unfounded claim that the election was stolen. Reports suggest that this claim was understood by some of the rioters as legitimizing their actions. The evidence here shows that Mr. Trump used the communicative authority of the presidency in support of attackers on the Capitol and an attempt to prevent the lawful counting of electoral votes.
- **Extent and reach:** Mr. Trump had a large audience, with a following of at least 35 million accounts on Facebook and at least 24 million accounts on Instagram.

million accounts on Facebook and at least 24 million accounts on Instagram.

Importantly, these social media posts are frequently picked up and shared more

broadly through mass media channels as well as by high-profile supporters of Mr. Trump with large audiences, greatly increasing their reach.

- **Imminence of harm:** The posts were made during a dynamic and fluid period of ongoing violence. There was a clear immediate risk of harm to life, electoral integrity, and political participation. The violence at the Capitol started within an hour of the rally organized through the use of Facebook and other social media. Indeed, even as Mr. Trump was posting, the rioters were rampaging through the halls of Congress and Members of Congress were expressing fear by calling on the White House and pleading for the president to calm the situation. The riot directly interfered with Congress's ability to discharge its constitutional responsibility of counting electoral votes, delaying this process by several hours.

Analyzing these factors, the Board concludes that the violation in this case was severe in terms of its human rights harms. Facebook's imposition of account-level restrictions on January 6 and the extension of those restrictions on January 7 was necessary and proportionate.

For the minority of the Board, while a suspension of an extended duration or permanent disablement could be justified on the basis of the January 6 events alone, the proportionality analysis should also be informed by Mr. Trump's use of Facebook's platforms prior to the November 2020 presidential election. In particular, the minority noted the May 28, 2020, post "when the looting starts, the shooting starts," made in the context of protests for racial justice, as well as multiple posts referencing the "China Virus." Facebook has made commitments to respect the right to non-discrimination (Article 2, para. 1 ICCPR, Article 2 ICERD) and, in line with the requirements for restrictions on the right to freedom of expression (Article 19, para. 3 ICCPR), to prevent the use of its platforms for advocacy of racial or national hatred constituting incitement to hostility, discrimination or violence (Article 20 ICCPR, Article 4 ICERD). The frequency, quantity and extent of harmful communications should inform the Rabat incitement analysis (Rabat Plan of Action, para. 29), in particular the factors on context and intent. For the minority, this broader analysis would be crucial to inform Facebook's assessment of a proportionate penalty on January 7, which should serve as both a deterrent to other political leaders and, where appropriate, an opportunity of rehabilitation. Further, if Facebook opted to impose a time-limited

suspension, the risk-analysis required prior to reinstatement should also take into account these factors. Having dealt with this case on other grounds, the majority does not comment on these matters.

9. Oversight Board decision

On January 6, Facebook's decision to impose restrictions on Mr. Trump's accounts was justified. The posts in question violated the rules of Facebook and Instagram that prohibit support or praise of violating events, including the riot that was then underway at the U.S. Capitol. Given the seriousness of the violations and the ongoing risk of violence, Facebook was justified in imposing account-level restrictions and extending those restrictions on January 7.

However, it was not appropriate for Facebook to impose an indefinite suspension.

Facebook did not follow a clear published procedure in this case. Facebook's normal account-level penalties for violations of its rules are to impose either a time-limited suspension or to permanently disable the user's account. The Board finds that it is not permissible for Facebook to keep a user off the platform for an undefined period, with no criteria for when or whether the account will be restored.

It is Facebook's role to create and communicate necessary and proportionate penalties that it applies in response to severe violations of its content policies. The Board's role is to ensure that Facebook's rules and processes are consistent with its content policies, its values, and its commitment to respect human rights. In applying an indeterminate and standardless penalty and then referring this case to the Board to resolve, Facebook seeks to avoid its responsibilities. The Board declines Facebook's request and insists that Facebook apply and justify a defined penalty.

Facebook must, within six months of this decision, reexamine the arbitrary penalty it imposed on January 7 and decide the appropriate penalty. This penalty must be based on the gravity of the violation and the prospect of future harm. It must also be consistent with Facebook's rules for severe violations which must in turn be clear, necessary, and proportionate.

If Facebook determines that Mr. Trump's accounts should be restored, Facebook should apply its rules to that decision, including any modifications made pursuant to the policy recommendations below. Also, if Facebook determines to return him to the

platform, it must address any further violations promptly and in accordance with its established content policies.

A minority believes that it is important to outline some minimum criteria that reflect the Board's assessment of Facebook's human rights responsibilities. The majority prefers instead to provide this guidance as a policy recommendation. The minority explicitly notes that Facebook's responsibilities to respect human rights include facilitating the remediation of adverse human rights impacts it has contributed to (UNGPs, Principle 22). Remedy is a fundamental component of the UNGP 'Protect, Respect, Remedy' framework, reflecting international human rights law more broadly (Article 2, para. 1, ICCPR, as interpreted by the Human Rights Committee in General Comment No. 31, paras. 15 - 18). To fulfil its responsibility to guarantee that the adverse impacts are not repeated, Facebook must assess whether reinstating Mr. Trump's accounts would pose a serious risk of inciting imminent discrimination, violence or other lawless action. This assessment of risk should be based on the considerations the Board detailed in the analysis of necessity and proportionality in Section 8.3.III above, including context and conditions on and off Facebook and Instagram. Facebook should, for example, be satisfied that Mr. Trump has ceased making unfounded claims about election fraud in the manner that justified suspension on January 6. Facebook's enforcement procedures aim to be rehabilitative, and the minority believes that this aim accords well with the principle of satisfaction in human rights law. A minority of the Board emphasizes that Facebook's rules should ensure that users who seek reinstatement after suspension recognize their wrongdoing and commit to observing the rules in the future. In this case, the minority suggests that, before Mr. Trump's account can be restored, Facebook must also aim to ensure the withdrawal of praise or support for those involved in the riots.

10. Policy advisory statement

The Board acknowledges the difficult issues raised by this case and is grateful for the many thoughtful and engaged public comments that it received.

In its referral of this matter to the Oversight Board, Facebook specifically requested "observations or recommendations from the board about suspensions when the user is a political leader." The Board asked Facebook to clarify their understanding of the term "political leader;" Facebook explained that they sought to cover "elected or appointed government officials and people who are actively running for office in an upcoming

election, including a short period of time after the election if the candidate is not elected” but not all state actors. Based on the Board’s analysis of this case, it confines its guidance to issues of public safety.

The Board believes that it is not always useful to draw a firm distinction between political leaders and other influential users. It is important to recognize that other users with large audiences can also contribute to serious risks of harm. The same rules should apply to all users of the platform; but context matters when assessing issues of causality and the probability and imminence of harm. What is important is the degree of influence that a user has over other users.

When posts by influential users pose a high probability of imminent harm, as assessed under international human rights standards, Facebook should take action to enforce its rules quickly. Facebook must assess posts by influential users in context according to the way they are likely to be understood, even if their incendiary message is couched in language designed to avoid responsibility, such as superficial encouragement to act peacefully or lawfully. Facebook used the six contextual factors in the Rabat Plan of Action in this case, and the Board thinks this is a useful way to assess the contextual risks of potentially harmful speech. The Board stresses that time is of the essence in such situations; taking action before influential users can cause significant harm should take priority over newsworthiness and other values of political communication.

While all users should be held to the same content policies, there are unique factors that must be considered in assessing the speech of political leaders. Heads of state and other high officials of government can have a greater power to cause harm than other people. Facebook should recognize that posts by heads of state and other high officials of government can carry a heightened risk of encouraging, legitimizing, or inciting violence - either because their high position of trust imbues their words with greater force and credibility or because their followers may infer they can act with impunity. At the same time, it is important to protect the rights of people to hear political speech. Nonetheless, if the head of state or high government official has repeatedly posted messages that pose a risk of harm under international human rights norms, Facebook should suspend the account for a determinate period sufficient to protect against imminent harm. Periods of suspension should be long enough to deter misconduct and may, in appropriate cases, include account or page deletion.

Restrictions on speech are often imposed by or at the behest of powerful state actors against dissenting voices and members of political oppositions. Facebook must resist

pressure from governments to silence their political opposition. When assessing potential risks, Facebook should be particularly careful to consider the relevant political context. In evaluating political speech from highly influential users, Facebook should rapidly escalate the content moderation process to specialized staff who are familiar with the linguistic and political context and insulated from political and economic interference and undue influence. This analysis should examine the conduct of highly influential users on the Facebook and Instagram platforms to adequately assess the full relevant context of potentially harmful speech. Further, Facebook should ensure that it dedicates adequate resourcing and expertise to assess risks of harm from influential accounts globally.

Facebook should publicly explain the rules that it uses when it imposes account-level sanctions against influential users. These rules should ensure that when Facebook imposes a time-limited suspension on the account of an influential user to reduce the risk of significant harm, it will assess whether the risk has receded before the suspension term expires. If Facebook identifies that the user poses a serious risk of inciting imminent violence, discrimination, or other lawless action at that time, another time-bound suspension should be imposed when such measures are necessary to protect public safety and proportionate to the risk.

When Facebook implements special procedures that apply to influential users, these should be well documented. It was unclear whether Facebook applied different standards in this case, and the Board heard many concerns about the potential application of the newsworthiness allowance. It is important that Facebook address this lack of transparency and the confusion it has caused. Facebook should produce more information to help users understand and evaluate the process and criteria for applying the newsworthiness allowance. Facebook should clearly explain how the newsworthiness allowance applies to influential accounts, including political leaders and other public figures. In regard to cross check review, Facebook should clearly explain the rationale, standards, and processes of review, including the criteria to determine which pages and accounts are selected for inclusion. Facebook should report on the relative error rates and thematic consistency of determinations made through the cross check process compared with ordinary enforcement procedures.

When Facebook's platform has been abused by influential users in a way that results in serious adverse human rights impacts, it should conduct a thorough investigation into the incident. Facebook should assess what influence it had and assess what changes it

could enact to identify, prevent, mitigate, and account for adverse impacts in future. In relation to this case, Facebook should undertake a comprehensive review of its potential contribution to the narrative of electoral fraud and the exacerbated tensions that culminated in the violence in the United States on January 6, 2021. This should be an open reflection on the design and policy choices that Facebook has made that may enable its platform to be abused. Facebook should carry out this due diligence, implement a plan to act upon its findings, and communicate openly about how it addresses adverse human rights impacts it was involved with.

In cases where Facebook or Instagram users may have engaged in atrocity crimes or grave human rights violations, as well as incitement under Article 20 of the ICCPR, the removal of content and disabling of accounts, while potentially reducing the risk of harm, may also undermine accountability efforts, including by removing evidence. Facebook has a responsibility to collect, preserve and, where appropriate, share information to assist in the investigation and potential prosecution of grave violations of international criminal, human rights and humanitarian law by competent authorities and accountability mechanisms. Facebook's corporate human rights policy should make clear the protocols the company has in place in this regard. The policy should also make clear how information previously public on the platform can be made available to researchers conducting investigations that conform with international standards and applicable data protection laws.

This case highlights further deficiencies in Facebook's policies that it should address. In particular, the Board finds that Facebook's penalty system is not sufficiently clear to users and does not provide adequate guidance to regulate Facebook's exercise of discretion. Facebook should explain in its Community Standards and Guidelines its strikes and penalties process for restricting profiles, pages, groups and accounts on Facebook and Instagram in a clear, comprehensive, and accessible manner. These policies should provide users with sufficient information to understand when strikes are imposed (including any applicable exceptions or allowances) and how penalties are calculated. Facebook should also provide users with accessible information on how many violations, strikes, and penalties have been assessed against them, as well as the consequences that will follow future violations. 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.

Finally, the Board urges Facebook to develop and publish a policy that governs its response to crises or novel situations where its regular processes would not prevent or avoid imminent harm. While these situations cannot always be anticipated, Facebook's guidance should set appropriate parameters for such actions, including a requirement to review its decision within a fixed time.

***Procedural note:**

The Oversight Board's decisions are prepared by panels of five Members and approved by a majority of the Board. Board decisions do not necessarily represent the personal views of all Members.

Data di pubblicazione

5 mag 2021

Piattaforma

Facebook

Standard della community pertinenti

Persone e organizzazioni pericolose

Argomenti correlati

Libertà di espressione, Politica, Sicurezza

Regione

Stati Uniti e Canada

Paesi interessati

Stati Uniti

Normative citate

Persone e organizzazioni pericolose