Brazilian general's speech (2023-001-FB-UA)

Case summary

The Oversight Board has overturned Meta's original decision to leave up a Facebook video which features a Brazilian general calling people to "hit the streets," and "go to the National Congress and the Supreme Court." Though the Board acknowledges that Meta set up several risk evaluation and mitigation measures during and after the elections, given the potential risk of its platforms being used to incite violence in the context of elections, Meta should continuously increase its efforts to prevent, mitigate and address adverse outcomes. The Board recommends that Meta develop a framework for evaluating its election integrity efforts to prevent its platforms from being used to promote political violence.

About the case

Brazil's presidential elections in October 2022 were highly polarized, with widespread and coordinated online and offline claims questioning the legitimacy of elections. These included calls for military intervention, and for the invasion of government buildings to stop the transition to a new government. The heightened risk of political violence did not subside with the assumption of office by newly elected President Luiz Inácio Lula da Silva on January 1, 2023, as civil unrest, protests, and encampments in front of military bases were ongoing.

Two days later, on January 3, 2023, a Facebook user posted a video related to the 2022 Brazilian elections. The caption in Portuguese includes a call to "besiege" Brazil's Congress as "the last alternative." The video also shows part of a speech given by a prominent Brazilian general who supports the re-election of former President Jair Bolsonaro. In the video, the uniformed general calls for people to

"hit the streets" and "go to the National Congress … [and the] Supreme Court." A sequence of images follows, including one of a fire raging in the Three Powers Plaza in Brasília, which houses Brazil's presidential offices, Congress, and Supreme Court. Text overlaying the image reads, in Portuguese, "Come to Brasília! Let's Storm it! Let's besiege the three powers." Text overlaying another image reads "we demand the source code," a slogan that protestors have used to question the reliability of Brazil's electronic voting machines.

On the day the content was posted, a user reported it for violating Meta's Violence and Incitement Community Standard, which prohibits calls for forcible entry into high-risk locations. In total, four users reported the content seven times between January 3 and 4. Following the first report, the content was reviewed by a content reviewer and found not to violate Meta's policies. The user appealed the decision, but it was upheld by a second content reviewer. The next day, the other six reports were reviewed by five different moderators, all of whom found that the content did not violate Meta's policies.

On January 8, supporters of former president Bolsonaro broke into the National Congress, Supreme Court, and presidential offices located in the "Three Powers Plaza" in Brasília, intimidating the police and destroying property. On January 9, Meta declared the January 8 rioting a "violating event" under its Dangerous Individuals and Organizations policy and said it would remove "content that supports or praises these actions." The company also announced that it had "designated Brazil as a Temporary High-Risk Location" and had "been removing content calling for people to take up arms or forcibly invade Congress, the Presidential palace and other federal buildings."

As a result of the Board selecting this case, Meta determined that its repeated decisions to leave the content on Facebook were in error. On January 20, 2023, after the Board shortlisted this case, Meta removed the content.

Key findings

This case raises concerns around the effectiveness of Meta's election integrity efforts in the context of Brazil's 2022 General Election, and elsewhere. While challenging the integrity of elections is generally considered protected speech, in some circumstances widespread claims which attempt to undermine elections can lead to violence. In this case, the speaker's intent, the content of the speech and its reach, as well as the likelihood of imminent harm resulting in the political context of Brazil at the time, all justified removing the post.

For a post to violate Meta's rules on calling for forcible entry into high-risk locations, the location must be considered "high-risk," and it must be situated in an area or vicinity that is separately designated as a "temporary high-risk location." As the post was an unambiguous call to forcibly enter government buildings situated in the Three Powers Plaza in Brasília ("high-risk locations" situated in a "temporary high-risk location," Brazil), Meta's initial decisions to leave this content up during a time of heightened political violence represented a clear departure from its own rules.

The Board is deeply concerned that despite the civil unrest in Brazil at the time the content was posted, and the widespread proliferation of similar content in the weeks and months ahead of the January 8 riots, Meta's content moderators repeatedly assessed this content as non-violating and failed to escalate it for further review. In addition, when the Board asked Meta for information on specific election-related claims on its platforms before, during, and after the Brazilian elections, the company explained that it does not have data on the prevalence of such claims. The content in this case was finally removed more than two weeks later, by which point the violating event it called had already occurred, and only after the Board brought the case to Meta's attention.

In response to a question from the Board, Meta said that it does not adopt any particular metrics for measuring the success of its election integrity efforts generally. Therefore, the Board finds that Meta should develop a framework for evaluating the company's election integrity efforts, and for public reporting on the subject. This aims to provide the company with relevant data to improve its content moderation system as a whole and to decide how best to employ its resources in electoral contexts. Without this kind of information, neither the Board nor the public can evaluate the effectiveness of Meta's election integrity efforts more broadly.

The Oversight Board's decision

The Oversight Board overturns Meta's original decision to leave up the post.

The Board also recommends that Meta:

- Develop a framework for evaluating its election integrity efforts. This includes creating and sharing metrics for successful election integrity efforts, including those related to Meta's enforcement of its content policies and its approach to ads.

- Clarify in its Transparency Center that, in addition to the Crisis Policy Protocol, the company runs other protocols in its attempt to prevent and address potential risk of harm arising in electoral contexts or other high-risk events.

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

Full case decision

1. Decision summary

1. The Oversight Board overturns Meta's original decision to leave up a Facebook video featuring a Brazilian general calling people to "hit the streets," and "go to the National Congress and the Supreme Court." These calls were followed by an image of the Three Powers Plaza in Brasília, where these government buildings are located, on fire, with overlay text which reads "Come to Brasília! Let's storm it! Let's besiege the three powers." The Board finds these statements to be clear and unambiguous calls to invade and take control of these buildings in the context of Bolsonaro supporters disputing election results and calling for military intervention to stop the course of a government transition. After the Board shortlisted this post for review, Meta reversed its original decision and removed it from Facebook.

2. The case raises broader concerns around the effectiveness of Meta's election integrity efforts in the context of Brazil's 2022 General Election, and elsewhere. Challenging elections' integrity is generally considered protected speech, but in some circumstances, widespread online and offline claims attempting to undermine elections, such as the ones in this case, can lead to offline violence. In Brazil, every warning signal was present that such violence would result. Though the Board acknowledges that Meta set up several risk evaluation and mitigation measures during and after the elections, given the potential risk of its platforms being used to incite violence in the context of elections, Meta should continuously increase its efforts to prevent, mitigate and address adverse outcomes. The post-election phase should be covered by Meta's election integrity efforts to address the risk of violence in a context of transition of power.

3. The Board therefore recommends that Meta develop a framework for evaluating the company's election integrity efforts and for public reporting on the subject. Such a framework should include metrics of success on the most relevant aspects of Meta's election integrity efforts, allowing the company not only to identify and reverse errors, but also to keep track of how effective its measures are in critical situations. The Board also recommends that Meta provides clarity in regards to the different protocols and measures it has in place to prevent and address potential risk of harm arising in

electoral contexts and other high-risk events. This includes naming and describing such protocols, their objective, the points of contact between them and how they differ from each other. Such protocols need to be more effective, have a clear chain of command, and be adequately staffed, especially when operating in a context of elections with a heightened risk of political violence. These recommendations would help improve the company's content moderation system as a whole by placing Meta in a better position to prevent its platforms from being used to promote political violence and to enhance its responses to election-related violence more generally.

2. Case description and background

4. On January 3, 2023, a Facebook user posted a video related to the 2022 Brazilian elections. The caption in Portuguese includes a call to "besiege" Brazil's Congress as "the last alternative." The one minute and 32-second video shows part of a speech given by a prominent Brazilian general and supporter of the reelection of former President Jair Bolsonaro. In the video, the uniformed general calls for people to "hit the streets" and "go to the National Congress … [and the] Supreme Court." A sequence of images follows, including one of a fire raging in the Three Powers Plaza in Brasília, which houses Brazil's presidential offices, Congress, and Supreme Court. Text overlaying the image reads, in Portuguese, "Come to Brasília! Let's Storm it! Let's besiege the three powers." Text overlaying another image reads "we demand the source code," a slogan that protestors have used to question the reliability of Brazil's electronic voting machines. The video was played over 18,000 times and was not shared.

5. Two days before the content was posted, Bolsonaro's electoral opponent Luiz Inácio Lula da Silva had been sworn-in as Brazil's president after winning the presidential run-off election on October 30, 2022 with 50.9 percent of the votes. The periods before, between, and after the two rounds of voting were marked by a heightened risk of political violence, spurred by claims about impending electoral fraud. This was premised on the alleged vulnerability of Brazil's electronic voting machines to hacking. Ahead of the election, then-President Bolsonaro fueled distrust in the electoral system, alleging fraud without supporting evidence and claiming that the electronic voting machines are not reliable. Some military officials echoed similar claims of electoral fraud and spoke in favor of using the military as an arbiter in electoral disputes. Several instances of political ads attacking the legitimacy of the elections on Meta's platforms were reported. These included posts and videos attacking judicial authorities and promoting a military coup. Further, Global Witness published a report on Brazil describing how political ads which violated the Community Standards were approved by the company and circulated on Meta's platforms. The findings tracked similar reports from the organization concerning other countries such as Myanmar and Kenya.

6. The post-election period was accompanied by civil unrest, including protests, roadblocks, and setting up encampments in front of military bases to call on the armed forces to overturn the election results. According to experts consulted by the Board, the video in this case first surfaced online in October 2022, soon after the electoral results were known; similar content remained on different social media platforms leading up to the January 8 riots. On December 12, 2022, the same day Lula's victory was confirmed by the Superior Electoral Court, a group of pro-Bolsonaro protesters tried to break into the headquarters of the Federal Police in Brasília. Several acts of vandalism took place. On December 24, 2022, there was an attempted bombing near the country's international airport in Brasília. The man responsible for the attack was arrested and confessed that his goal was to attract attention to their pro-coup cause.

7. The heightened risk of political violence in Brazil did not subside with the newly elected president's inauguration on January 1, 2023. Based on research commissioned by the Board, false claims about voting machines peaked on Meta platforms after the first and second rounds of voting, and again in the weeks following Lula's victory. Additionally, in the days leading up to January 8, Bolsonaro supporters used several coded slogans to promote protests in Brasília which were specifically focused on government buildings. Most of the logistical organization appeared to be accomplished through communications channels other than Facebook.

8. International election observation missions such as the Organization of American States and the Carter Center reported that there was no substantial evidence of fraud and that the election had been conducted in a free and fair manner despite the pressures of a highly polarized electorate. The Brazilian Ministry of Defense also formally observed the election and reported no evidence of irregularities or fraud, though it did subsequently release a conflicting statement that the armed forces "do not rule out the possibility of fraud." In Brazil, the Ministry of Defense oversees the work of the armed forces.

9. Tensions culminated on January 8, when supporters of former President Bolsonaro broke into the National Congress, Supreme Court, and presidential offices located in the "Three Powers Plaza" in Brasília referred to in the case content, intimidating the police and destroying property. Around 1,400 people were arrested for participating in the January 8 riots, with around 600 still in custody.

10. In the wake of the events of January 8, the United Nations condemned the use of violence, saying that it was the "culmination of the sustained distortion of facts and incitement to violence and hatred by political, social and economic actors who have been fueling an atmosphere of distrust, division, and destruction by rejecting the result of democratic elections." It reiterated its commitment and confidence in Brazil's democratic institutions. Public comments and experts consulted by the Board

indicated the harmful effect that claims which preemptively cast doubt on the integrity of Brazil's electoral system had in driving political polarization and enabling offline political violence (See public comments from the Dangerous Speech Project [PC-11010], LARDEM - Clínica de Direitos Humanos da Pontifícia Universidade Católica do Paraná [PC-11011], Instituto Vero [PC-11015], ModeraLab [PC-11016], Campaign Legal Center [PC-11017], Center for Democracy & Technology [PC-11018], InternetLab [PC-11019], and Coalizão Direitos na Rede [PC-11020]).

11. On January 9, 2023, Meta declared the January 8 rioting a "violating event" under the Dangerous Individuals and Organizations policy and said it would remove "content that supports or praises these actions." The company also announced that "[i]n advance of the election" it had "designated Brazil as a Temporary High-Risk Location" and had "been removing content calling for people to take up arms or forcibly invade Congress, the Presidential palace and other federal buildings."

12. On January 3, the same day the content was posted, a user reported it for violating Meta's Violence and Incitement Community Standard, which prohibits calls to "forcibly enter locations . . . where there are temporary signals of a heightened risk of violence or offline harm." In total, four users reported the content seven times between January 3 and 4. Following the first report, the content was reviewed by a human moderator and found not to violate Meta's policies. The user appealed the decision, but it was upheld by a second human moderator. The next day, the other six reports were reviewed by five different moderators, all of whom found that the content did not violate Meta's policies. The content was not escalated to policy or subject matter experts for additional review. In response to a question from the Board, Meta clarified that the seven people who reviewed the content were based in Europe. According to Meta, they were all fluent in Portuguese and had the language and cultural expertise to review Brazilian content.

13. As a result of the Board selecting this case, Meta determined that its repeated decisions to leave the content on Facebook were in error. On January 20, 2023, after the Board shortlisted the case, Meta removed the content, issued a strike against the content creator's account, and applied a 24-hour feature-limit, preventing them from creating new content within that period. Despite Meta's action, civil society group Ekō's public comment submission to the Board and other reports emphasized that similar content remained on Facebook even after this case was brought to Meta's attention by the Board (PC-11000).

3. Oversight Board authority and scope

14. The Board has authority to review Meta's decision following an appeal from the person who previously reported content that was left up (Charter Article 2, Section 1; Bylaws Article 3, Section 1). The Board may uphold or overturn Meta's decision

(Charter Article 3, Section 5), and this decision is binding on the company (Charter Article 4). Meta must also assess the feasibility of applying its decision in respect of identical content with parallel context (Charter Article 4). The Board's decisions may include non-binding recommendations that Meta must respond to (Charter Article 3, Section 4; Article 4). Where Meta commits to act on recommendations, the Board monitors their implementation.

15. When the Board selects cases like this one, where Meta subsequently acknowledges that it made a mistake, the Board reviews the original decision, to help increase understanding of the policy parameters and content moderation processes that contributed to the error. The Board then seeks to address issues it identifies with Meta's underlying policies or processes. The Board also aims to issue recommendations for Meta to improve enforcement accuracy and treat users fairly moving forward.

4. Sources of authority and guidance

16. The following standards and precedents informed the Board's analysis in this case:

I. *Oversight Board decisions:*

17. The most relevant previous decisions of the Oversight Board include:

- "[Former President Trump's suspension" (case decision 2021-001-FB-FBR):](#) The Board noted that, in electoral contexts, Meta's human rights responsibilities require allowing political expression while avoiding serious risks to other human rights.
- "[Myanmar bot" (case decision 2021-007-FB-UA):](#) The Board highlighted the importance of protecting political speech during periods of political crisis.
- "[Tigray Communication Affairs Bureau" (case decision 2022-006-FB-MR):](#) The Board highlighted Meta's responsibility to establish a principled and transparent system for moderating content in conflict zones to mitigate the risks of its platforms being used to incite violence.
- "[Knin cartoon" (case decision 2022-001-FB-UA):](#) The Board urged Meta to provide more clarity on how content gets escalated to subject matter experts.

II. *Meta's content policies:*

Violence and Incitement Community Standard

18. Under the Violence and Incitement Community Standard, Meta does not permit "statements of intent or advocacy, calls to action, or aspirational or conditional statements to forcibly enter locations (including but not limited to places of worship, educational facilities, polling places or locations used to count votes or administer an election) where there are temporary signals of a heightened risk of violence or offline harm." The policy rationale for this Community Standard is to "prevent potential offline harm that may be related to content" appearing on Meta's platforms. At the same time, Meta recognizes that "people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways." Meta therefore removes content when the company believes "there is a genuine risk of physical harm or direct threats to public safety." In determining whether a threat is credible, Meta also considers "the language and the context."

19. The Board's analysis was also informed by Meta's commitment to "Voice," which the company describes as "paramount," and its value of "Safety."

III. Meta's human rights responsibilities

20. The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of companies. In 2021, Meta announced its Corporate Human Rights Policy, where it reaffirmed its commitment to respecting human rights in accordance with the UNGPs.

21. The Board's analysis of Meta's human rights responsibilities in this case was informed by the following international standards:
- The right to freedom of opinion and expression: Articles 19 and 20, International Covenant on Civil and Political Rights (ICCPR), General Comment No. 34, Human Rights Committee, 2011; Research Paper 1/2019 on Elections in the Digital Age (2019): UN Special Rapporteur on freedom of opinion and expression, reports: A/HRC/38/35 (2018) and A/74/486 (2019); Rabat Plan of Action, UN High Commissioner for Human Rights report: A/HRC/22/17/Add.4 (2013).
- The right to peaceful assembly: Article 21, ICCPR; General Comment No. 37, Human Rights Committee, 2020.
- The right to life: Article 6, ICCPR.
- The right to participate in public affairs and the right to vote: Article 25, ICCPR.

5. User submissions

22. In their appeal to the Board, the user who reported the content stated that they "have already reported this and countless other videos to Facebook and the answer is always the same, that it doesn't violate the Community Standards." The user further linked the content's potential to incite violence to action taken by people in Brazil "who do not accept the results of elections."

6. Meta's submissions

23. When the Board brought this case to Meta's attention, the company determined that its original decision to leave the content up was incorrect. Meta provided the Board with a broad analysis of Brazil's social and political context before, during, and after the presidential election to justify the – albeit belated – removal of the content in this case. It later provided the Board with probable factors that "may have contributed" to the persistent enforcement error.

24. Meta stated its view that "the multiple references to 'besieging' high-risk locations in the caption and video do not independently rise to the level of 'forcible entry' under [the] [Violence and Incitement] policy." However, "the combination of calling on people to 'Come to Brasília! Let's storm it! Let's besiege the three powers' with the background image of the Three Powers Plaza on fire makes the intent to forcibly enter these prominent locations clear."

25. According to Meta, the content did not qualify for a newsworthiness allowance even though it acknowledged that its platforms are "important places for political discourse, especially around elections." In this case, the public interest value of the content did not outweigh the risk of harm given its "explicit call for violence" and the "heightened risk of offline harm following the Brazilian Presidential election and Lula's inauguration." Meta found no indication that the content was shared to condemn or raise awareness of the call for violence. The company maintains that its ultimate decision to remove the content is consistent with its values and with international human rights standards.

26. To address elections and other crisis situations, Meta has set up several risk evaluation and mitigation measures that are run by different teams and can apply either simultaneously or independently. Each has different "tiers" or "levels" of intensity depending on the respective risk evaluation:

- The [Integrity Country Prioritization policy](a.k.a., at-risk tiering system), which is run by the product team within Meta, provides a framework for the long-term prioritization of product resource investments. While Meta describes this

11

process as unresponsive to short-term crises, it does evaluate all countries twice per year for emerging risks/threats.

- The [Integrity Product Operations Center](#) (IPOCs), brings together a cross-functional team of subject matter experts from across the company to "respond in real time to potential problems and trends." IPOCs are set up to quickly assess a large set of issues, identify risks, and determine how to address them in the context of a crisis or high-risk situation. IPOCs are called Election Operation Centers when they are specifically focused on elections.
- [Election Operation Centers](#) offer "real-time monitoring on key elections issues, such as efforts to prevent people from voting, increases in spam, potential foreign interference, or reports of content that violates [Meta's] policies", and "monitor news coverage and election-related activity across other social networks and traditional media." The centers provide Meta with a "collective view and help track what type of content may go viral" to "accelerate" the company's response time to these threats. Part of Election Operation Center preparation involves "extensive scenario-planning to game out potential threats – from harassment to voter suppression – and develop systems and procedures in advance to respond effectively."
- Finally, the [Crisis Policy Protocol](#) is the framework Meta adopted for developing time-bound policy-specific responses to an emerging crisis. Meta developed this protocol in response to an Oversight Board recommendation in the [Former President Trump's suspension case](#). Under this protocol, Meta establishes [three crisis categories](#), based on which the company adopts a given set of measures to mitigate risks. A Category 1 crisis is, for instance, triggered by "increased law enforcement or military activity" or a "planned high-risk election or flashpoint event."

27. The Election Operation Center covering the 2022 Brazilian general election ran at various points in time from September to November 2022, including during the first and second rounds of the election. However, there was no Election Operation Center (or IPOC) in place at the time the content was posted on January 3, 2023. Meta designated the "post-election unrest" as a crisis under the Crisis Policy Protocol to help the company assess how best to mitigate content risks.

28. In response to a question from the Board regarding digital trends on Meta's platforms before, during and after the Brazilian elections, the company stated that as part of its "election preparation and response work, a number of teams identified election-related content trends and incorporated them into [their] risk-mitigation

strategy." These included: "(i) risks associated with incitement or spread of threats of violence; (ii) misinformation; and (iii) business integrity, which include risks associated with potential abuse of advertisement with harmful content... or attempts to conduct campaigns in ways that manipulate or corrupt public debate." Meta stated that the "results, among other factors, helped inform a number of product and policy mitigations." However, Meta does not have "prevalence data" on specific claims (e.g. of electoral fraud, calls to go to Brasília or forcibly invade federal government buildings, calls for a military intervention), because in general, the company's enforcement systems "are set up to monitor and track based on the policies they violate."

29. The Board asked Meta 15 questions in writing, including 5 in follow-up to an oral briefing on how Election Operation Centers work. Questions related to: policy levers available to address coordinated behavior on Meta's platforms; risks identified ahead of the 2022 Brazilian elections; the relationship between the Election Operation Center for the Brazilian election and the Crisis Policy Protocol; how Meta draws the line when distinguishing between legitimate political organizing and harmful coordinated action; digital trends on Meta's platforms in Brazil before, during, and after the elections; and the language capabilities of the content moderators who reviewed the case content.

30. Meta answered 13 questions. Meta did not answer two questions, one concerning the relationship between political advertising and misinformation, and another concerning the number of removals of pages and accounts while the Election Operation Center for the 2022 Brazil elections was in place. Meta also informed the Board that the company did not have more general data on content moderation in the context of Brazil's 2022 election readily available to share with the Board, in addition to the number of content takedowns which was already shared publicly. Meta further explained that the company does not assess its performance in the context of elections against a given set of metrics of success and benchmarks. Meta raised the need to prioritize resources when responding to the Board's questions, and said that providing the requested data within the timeframe for deciding the case would not be possible.

7. Public comments

31. The Oversight Board received 18 public comments relevant to this case. Eleven of the comments originated from Latin America and the Caribbean, three from the United States and Canada, two from the Middle East and North Africa, one from Asia Pacific and Oceania, and one from Central and South Asia. Additionally, in February 2023, the Board organized a roundtable with stakeholders from Brazil and Latin America on the topic of "Content Moderation and Political Transitions."

32. The submissions covered the following themes: the accumulation of harmful claims about election fraud and calls for a military coup on social media platforms before, during, and after the 2022 Brazil elections; election-related disinformation; Meta's election integrity efforts; Meta's responsibility to protect users' rights in the context of a democratic transition of power; the relationship between election denialism and political violence; and the importance of content reviewers' familiarity with the local political context.

33. To read public comments submitted for this case, please click here.

8. Oversight Board analysis

34. The Board examined whether this content should be removed by analyzing Meta's content policies, human rights responsibilities, and values. This case was selected because it allows the Board to assess how Meta distinguishes peaceful organizing on its platforms from incitement or coordination of violent action, especially in a context of a transition of power. Additionally, the case allows the Board to examine Meta's election integrity efforts more generally, and in Brazil more specifically, considering that post-election periods are crucial moments both to contest the integrity of an election and to guarantee that legitimate electoral results are respected. Therefore, the Board finds that Meta's election integrity efforts should cover both the electoral process itself and the post-electoral period, for the latter is also vulnerable to manipulation, election-related misinformation, and threats of violence. The case falls within the Board's "elections and civic space" strategic priority.

8.1 Compliance with Meta's content policies

I. Content rules

Violence and Incitement

35. The Board finds that the content in this case violates the Violence and Incitement Community Standard's prohibition of content calling for forcible entry into certain high-risk locations. The Board finds that while Meta's value of "Voice" is particularly relevant in electoral processes, including the post-electoral period, removing the content is necessary in this case to advance Meta's value of "Safety."

36. In order to violate the policy line against calls for forcible entry into high-risk locations, two "high-risk" designations are required. Firstly, the location must be considered "high-risk," and, secondly, it must be situated in an area or vicinity that is

separately designated as a Temporary High-Risk Location. Meta's specific instructions to content reviewers is to "[r]emove calls to action, statements of intent, statements advocating, and aspirational statements to forcibly enter high-risk locations within a Temporary High-Risk Location."

37. Meta defines a "high-risk location" as a "location, permanent or temporary, that is deemed high-risk due to its likelihood of being the target of violence." Permanent high-risk locations include "places of work or residence of high-risk persons or their families (for example, the headquarters for a news organization, medical centers, laboratories, police stations, government offices, etc.); facilities used during local, regional, and national elections as a voter registration center, polling location, vote counting site (for example local library, government building, community or civic center, etc.) or a site used in the administration of an election." According to Meta, the Brazilian Congress, Supreme Court and Presidential offices are all permanent "high-risk locations" by virtue of being places of work or residence of high-risk persons or their families."

38. The additional "Temporary High-Risk Location" designation of the broader area or vicinity covers any "location temporarily designated by [Meta as such] for a time-bound period." A place is designated as a Temporary High-Risk Location based on many factors, including "whether high-severity violence occurred at a protest in the location in the last 7 days;" "evidence of an increased risk of violence associated with civil unrest or a contentious court decision at the location;" "an assessment from law enforcement, internal security reports, or a trusted partner that imminent violence is likely to occur at the location;" "evidence of planned or active protest at the location or a planned or active protest at the location where the organizer has called for armaments to be used or brought to the location of the protest;" and "an assessment by internal teams that the safety concerns outweigh the potential impact on the expression of self-defense and self-determination." Once a Temporary High-Risk Location is designated, the designation is shared with Meta's internal teams. Though such designations are time-limited, the company occasionally grants extensions. According to Meta, a Temporary High-Risk Location designation leads to the proactive review of content "before users report [it]."

39. For the 2022 elections, Meta designated the entire country of Brazil as a Temporary High-Risk Location. The designation was initially established on September 1, 2022 based on Meta's assessment of increased risk of violence associated with ongoing civil and election-related unrest. The designation was extended to cover the October 2022 election and its aftermath, until February 22, 2023. The designation was in place at the time the case content was posted.

40. According to Meta, both designations must be present for a piece of content to violate the policy, which was the case for the post under analysis. According to Meta,

the two-fold requirement helps ensure that calls for protests are not broadly suppressed and that only content likely to result in violence will be removed.

41. Given the above, the Board regards Meta's initial decisions that the content should remain on the platform during a time of heightened risk of political violence as a clear departure from its own standard, because it constituted an unambiguous call to forcibly enter government buildings situated in the Three Powers Plaza in Brasília, which are "high-risk locations" situated in a "temporary high-risk location," Brazil.

II. Enforcement Action

42. According to Meta, seven human moderators who possessed the necessary linguistic and cultural expertise reviewed the content. Meta does not instruct at-scale reviewers to record their reasons for making decisions. When the Board selected this case, Meta's internal teams conducted an analysis which concluded that three probable factors "may have contributed" to the persistent enforcement error: (1) reviewers may have misunderstood the user's intent (a call to action) possibly due to a lack of punctuation that led to misinterpretation of the content as a neutral comment about the event; or (2) reviewers made a wrong decision despite the correct guidelines being in place due to multiple updates around the handling of content related to high-risk events from various sources; or (3) reviewers may not have seen the violation in the video.

43. Factors 1 and 3 suggest that moderators did not review this content carefully nor watched the video fully, as the potential violation of Meta's policies it contained was clear. However, Meta does not provide any explanation as to why the content was not escalated to subject matter and policy experts for further analysis. The content was not escalated despite the fact that it came from a country which, at the time the content was posted and reported, was designated as a "Temporary High-Risk Location" relating to a policy line that is only activated when this designation is in place. The content was also not escalated despite the overall online and offline context in Brazil (See Section 2).

44. Meta already [informed the Board](#) that content reviewers are not always able to watch videos in full. Nonetheless, in situations of heightened risk of violence, where specific policy levers have been triggered, the Board would expect content reviewers to be oriented to watch videos in full, as well as to escalate potentially violating content.

45. In relation to factor 2, while Meta stated it informs at-scale reviewers of Temporary High-Risk Location designations, the company acknowledges possible shortcomings in its socialization of this and other election-specific risk mitigation measures. The socialization of this kind of information enables content reviewers to detect, remove,

or escalate problematic content such as the video in this case. The fact that different evaluation and mitigation measures were in place in Brazil at the time indicates that they likely need to be better articulated and have a clearer chain of command to make the company's election integrity efforts more effective.

46. Despite Meta's ultimate decision to take down the content, the Board is deeply concerned that even with the civil unrest in Brazil at the time the content was posted, and the widespread proliferation of similar online content months and weeks before the January 8 riots, Meta's content moderators repeatedly assessed this content as non-violating and failed to escalate it for further review despite the contextual cues it contained. These concerns are compounded by the fact that when the Board asked Meta for information on specific election-related claims on its platforms before, during, and after the Brazilian election, the company explained it does not have such prevalence data (see Section 6). The content in this case was finally removed more than two weeks later, after the violating event it had called for already occurred, and only after the Board brought the case to Meta's attention.

47. Meta acknowledged the heightened risk of violence in Brazil, first by adopting various risk evaluation measures before, during, and after the content was posted, and also directly to the Board when the company decided to finally remove the content. Yet, the company's reviewers persistently failed to adequately enforce its Community Standards, particularly the very policy line of the Violence and Incitement Community Standard triggered by a Temporary High-Risk Location designation. The fact that the content was not escalated prior to Board selection, despite the clarity of the potential violation, and that there was similar content circulating on Facebook at the time (See Sections 2 and 8.2), indicates that escalation channels are likely to be insufficiently clear and effective (See Knin cartoon case). It also demonstrates the need for Meta to improve its safeguards around elections. As the Board has noted in previous decisions, it is indispensable that at-scale reviewers possess adequate linguistic and contextual knowledge and are equipped with the necessary tools and channels to escalate potentially violating content.

III. Transparency

48. The Board recognizes that Meta made important efforts to safeguard the integrity of the 2022 Brazil elections. In August 2022, when the campaign period formally began, Meta [publicly announced](#) its election-related initiatives in the country. The company worked with Brazil's Superior Electoral Court to add a label to posts about elections on Facebook and Instagram, "directing people to reliable information on the Electoral Justice website." According to Meta, this led to a "10-fold increase" in visits to the website. The partnership also allowed the Superior Electoral Court to report potentially violating content directly to Meta. Meta hosted training sessions for

electoral officials throughout Brazil to explain the company's Community Standards and how misinformation on Facebook and Instagram is addressed. Meta also prohibited paid advertising "calling into question the legitimacy of the upcoming election." Further, the company implemented a WhatsApp forwarding limit so that a message can only be forwarded to one WhatsApp group at a time. Finally, Meta reported the number of pieces of content removed under various Community Standards, such as the Violence and Incitement, Hate Speech, and Bullying and Harassment policies, and the total number of click-throughs on election labels that directed users to authoritative information about the Brazil elections.

49. Nonetheless, when asked by the Board about its election integrity efforts in the context of the 2022 Brazil elections, Meta stated that the company does not adopt any particular metrics for measuring the success of its election integrity efforts generally, beyond reporting data on content takedowns, views, and click-throughs on election labels. The Board also notes that, from Meta's disclosures in its Transparency Center and exchanges with the Board, it is not entirely clear how the company's different risk evaluation measures and protocols run (See Section 6 above), independently or in parallel. Meta should clarify the points of contact between these different protocols, better explain how they differ from each other and how exactly the enforcement of content policies is affected by them.

50. A number of public comments (Ekō [PC-11000], Dangerous Speech Project [PC-11010], ModeraLab [PC-11016], Campaign Legal Center [PC-11017], InternetLab [PC-11019], and Coalizão Direitos na Rede [PC-11020]) received by the Board stated that the company's efforts to safeguard the elections in Brazil were not sufficient. While the Board acknowledges the challenges inherent to moderating content at scale, Meta's responsibility to prevent, mitigate and address adverse human rights impacts is heightened in electoral and other high-risk contexts, and requires the company to establish effective guardrails against them. The enforcement error in this case does not appear to be an isolated incident. According to Ekō (PC-11000), similar content remained on Facebook even after the January 8 riots.

51. More transparency is needed to assess whether Meta's measures are adequate and sufficient throughout election contexts. The lack of data available for the Board to review undermined the Board's ability to adequately assess whether the enforcement errors in this case, and concerns raised by different stakeholders, are symptomatic of a systemic issue in the company's policies and enforcement practices. It also compromised the Board's ability to issue more specific recommendations for Meta on how to further improve its election integrity efforts globally.

52. Meta's current data disclosures, predominantly on content takedowns, do not give a complete picture of the outcome of the election integrity measures it puts in place in a given market. For instance, they do not include enforcement accuracy in relation to

important policies in electoral contexts, such as the Violence and Incitement Community Standard, nor the percentage of political ads initially approved by Meta but then found to violate its policies. Performing statistical auditing with metrics like these would allow Meta not only to reverse errors, but also to keep track of how effective its measures are when getting it right is of the utmost importance.

53. Without this kind of information, neither the Board nor the public can evaluate the effectiveness of Meta's election integrity efforts more broadly. This is important considering that many incidents of political violence often result from or are intensified by election-related disputes, where harmful content remained online to precede or accompany offline violence (See "Myanmar bot" (2021-007-FB-UA), "Tigray Communication Affairs Bureau" (2022-006-FB-MR), and "Former President Trump's suspension" (2021-001-FB-FBR)).

54. Therefore, the Board finds that Meta should develop a framework for evaluating the company's election integrity efforts, and for public reporting on the subject. This aims to provide the company with relevant data to improve its content moderation system as a whole and decide how to best employ its resources in electoral contexts. It should also help Meta to effectively draw on local knowledge and to identify and evaluate coordinated online and offline campaigns aimed at disrupting democratic processes. Additionally, this framework should be useful for Meta to set up permanent feedback channels, and to determine measures to be adopted when political violence persists after the formal conclusion of electoral processes. Finally, the Board notes that, as explained above, the articulation between Meta's different risk evaluation measures and protocols, such as the IPOCs, the Integrity Country Prioritization policy, and the Crisis Policy Protocol (See Section 6 above) in election-related contexts needs to be reviewed and better explained to the public.

8.2 Compliance with Meta's human rights responsibilities

1. *Freedom of expression (Article 19 ICCPR)*

55. The right to freedom of opinion and expression is a "central pillar of democratic societies, and a guarantor of free and fair electoral processes, and meaningful and representative public and political discourse" (UN Special Rapporteur on freedom of expression, Research Paper 1/2019, p. 2). Article 19 of the ICCPR provides for broad protection of expression, especially for political speech. Where restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimacy, and necessity and proportionality (Article 19, para. 3, ICCPR). These requirements are often referred to as the "three-part test." The Board uses this framework to interpret Meta's voluntary human rights commitments.

19

I. Legality (clarity and accessibility of the rules)

56. The principle of legality under international human rights law requires rules that limit expression to be clear and publicly accessible (General Comment No.34, at para. 25). Applied to the rules of social media companies, the UN Special Rapporteur on freedom of expression has said they should be clear and specific ([A/HRC/38/35](), para. 46). People using Meta's platforms should be able to access and understand the rules, and content reviewers should have clear guidance on their enforcement.

57. The Board finds that, as applied to the facts of this case, Meta's prohibition of content calling for forcible entry into certain high-risk locations is clearly stated, and the exact conditions under which the prohibition is triggered are likewise clear. The case content could be easily understood as violating both by the user and content reviewers, especially in Brazil's context of civil unrest. Therefore, the Board considers the legality requirement to be satisfied.

II. Legitimate aim

58. Restrictions on freedom of expression (Article 19, ICCPR) must pursue a legitimate aim. The Violence and Incitement policy aims to "prevent potential offline harm" by removing content that poses "a genuine risk of physical harm or direct threats to public safety." This policy serves the legitimate aim of protecting the rights of others, such as the right to life (Article 6, ICCPR), as well as public order and national security (Article 19, para. 3, ICCPR). In electoral contexts, this policy may also pursue the legitimate aim of protecting others' right to vote and participate in public affairs (Article 25, ICCPR).

III. Necessity and proportionality

59. The principle of necessity and proportionality provides that any restrictions on freedom of expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected" ([General Comment No. 34,]() para. 33 and 34). As in prior cases involving incitement to violence, the Board finds the six UN [Rabat Plan of Action]() factors relevant to determining the necessity and proportionality of the restriction (see, for example: [Former President Trump's suspension case]()).

60. The Board recognizes that in many political environments, challenging the integrity of the elections or the electoral system is a legitimate exercise of people's

rights to freedom of expression and protest, even if there are isolated incidents of violence. Due to their political message, they enjoy a heightened level of protection (General Comment No. 37, paras. 19 and 32). The Board notes, however, that this is not the case here. There is a crucial line distinguishing between protected political speech and incitement to violence to overturn the results of a lawful popular election. Based on the factors outlined in the Rabat Plan of Action, the threshold for speech restriction was clearly met in this case. The Board finds that several elements in the case content are relevant to its analysis: the calls to "besiege" Brazil's Congress as "the last alternative" and to "storm" the "three powers"; the video with a call from a prominent Brazilian general to "hit the streets" and "go to the National Congress … [and the] Supreme Court; the image of the federal government buildings burning in the background; and the demand for "the source code." They all are, in the wider Brazilian context of Bolsonaro supporters disputing the election results and asking for a military coup, an unambiguous call to invade and take control of government buildings. The intent of the speaker, the content of the speech and its reach, as well as the likelihood of imminent harm resulting in the political context of Brazil at that time, all justified removing the post.

61. The content was posted in a context of heightened risk of political violence, with widespread ongoing calls on the armed forces to overturn the election results. At the same time, coded slogans were being used to promote protests specifically focused on government buildings in Brasília (See Section 2). In this regard, information the Board received through several public comments, including from ITS Rio – Modera Lab (PC-11016), Coalizão Direitos na Rede (PC-11020), InternetLab (PC-11019), and Ekō (PC-11000), which supported research commissioned by the Board, all show that similar content was circulating widely on social media in the lead-up to the January 8 events. They also underscore the imminence of Bolsonaro supporters storming buildings at the Three Powers Plaza, and pushing the military to intervene, including through a military coup.

62. Given the above, the Board finds that the removal of the content is consistent with its human rights responsibilities. Removing the content is a necessary and proportionate response to protect the right to life of people, including public officials, and public order in Brazil. The removal of this and similar pieces of content is also necessary and proportionate to protect Brazilians' right to vote and participate in public affairs, in a context where attempts to undermine a democratic transition of power were underway.

63. The persistent failure of Meta's review systems to properly identify the violation in the video or escalate it for further review and remove the case content is a serious concern, which the Board believes Meta will be in a better position to address if the company implements the recommendations below. While Meta took positive steps to improve its election integrity efforts in Brazil, it has not done enough to address the potential misuse of its platforms through coordinated campaigns of the kind seen in

Brazil. In this case, the content that was left-up and widely shared appeared to be typical of the kind of misinformation and incitement reported to be circulating on Meta's platforms in Brazil at the time. It further substantiates claims that influential accounts with significant powers of mobilization on Meta's platforms had played a role in promoting violence. As asserted in public comments the Board received (See, Instituto Vero [PC-11015], ModeraLab [PC-11016], InternetLab [PC-11019], Instituto de Referência em Internet e Sociedade [PC-11021]), the review and potential removal of individual pieces of content from Meta's platforms is insufficient and relatively ineffective when such content is part of an organized and coordinated action aimed at disrupting democratic processes. Election integrity efforts and crisis protocols need to address these broader digital trends.

8.3 Identical content with parallel context

64. The Board expresses concern with the proliferation of content similar to the one under analysis in the months preceding the January 8 riots in Brazil. Given Meta's repeated failure in identifying this piece of content as violating, the Board will pay special attention to Meta's application of its decision to identical content with parallel context that has remained on the company's platforms, except when shared to condemn or raise awareness around the general's speech and the calls for storming the Three Powers Plaza buildings in Brasília.

9. Oversight Board decision

65. The Oversight Board overturns Meta's original decision to leave up the content.

10. Recommendations

A. Enforcement

1. Meta should develop a framework for evaluating the company's election integrity efforts. This includes creating and sharing metrics for successful election integrity efforts, including those related to Meta's enforcement of its content policies and the company's approach to ads. The Board will consider this recommendation implemented when Meta develops this framework (including a description of metrics and goals for those metrics), discloses it in the company's Transparency Center, starts publishing country-specific reports, and publicly discloses any changes to its general election integrity efforts as a result of this evaluation.

B. Transparency

2. Meta should clarify in its Transparency Center that, in addition to the Crisis Policy Protocol, the company runs other protocols in its attempt to prevent and address potential risk of harm arising in electoral contexts or other high-risk events. In addition to naming and describing those protocols, the company should also outline their objective, what the points of contact between these different protocols are, and how they differ from each other. The Board will consider this recommendation implemented when Meta publishes the information in its Transparency Center.

*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.

For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by an independent research institute headquartered at the University of Gothenburg which draws on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The Board was also assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. Memetica, an organization that engages in open-source research on social media trends, also provided analysis. Linguistic expertise was provided by Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world.