

23 March 2023

The Center for Democracy & Technology welcomes the opportunity to provide comments on case 2023-001-FB-UA, regarding a user’s appeal to remove another user’s post containing a video of a Brazilian general calling for supporters of the former Brazilian president to take over the National Congress in Brasília.

A user shared a video playing a portion of the General’s remarks where, through his speech and through overlaid text, viewers are called on to “besiege” Brazil’s congress and “demand the source code.” This video was reported seven times and reviewed by human reviewers numerous times. Each reviewer found that the post did not violate 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.” Upon selection of this case by the Oversight Board, Meta removed the post, saying that the post did in fact violate Meta’s policies. In this comment we explain why election integrity efforts must take into account the unique context of the country’s electoral process, be watchful of narratives that are disseminated before and during the electoral period and even after the transition of power, and be more transparent in their application and enforcement of content policies.

Background on the narratives circulating around the 2022 Brazilian election

The user posted on January 3, 2023, two days after the swearing in of President Luiz Inácio Lula da Silva and five days before the attack on the National Congress. At this point, supporters of the former President had already been camping outside the National Congress since his electoral loss; on December 12, supporters of former President Bolsonaro clashed with security forces in the nation’s capital.¹ In covering the attack on the Congress, Politico wrote that the remarks shared by supporters of the former president were an effort to “create chaos...rather than chang[e] the results of the election” as evinced by the fact that the video was posted after the inauguration of President Lula.²

A text overlay on the video reading “we demand the source code” references the electronic voting systems used to facilitate and tally the election results. This caption evokes the narrative, widespread among Bolsonaro’s supporters, that the voting system was rigged to cause Bolsonaro’s electoral loss. As our report on election disinformation in Brazil, the US, and France finds, this narrative was a common element in disinformation campaigns that have circulated across Brazil since 2018.³

This context is critical to underscore the unique state of Brazilian elections and to situate the post at the heart of this case within a longer lineage of speech intended to sow distrust and mobilize

¹ [Brazil's attack on Congress explained - CNN](#)

² [The American politics of Brazil's insurrection - POLITICO](#)

³ [Report - A Lie Can Travel: Election Disinformation in the United States, Brazil, and France](#)

the nation's citizens to reject the results of elections. A reviewer familiar with Brazil's elections may have been able to ascertain quickly the post's intended impact. Although Meta indicated that they were staffing up their local Brazil moderation team before the elections, we do not know the expertise of the six reviewers who reviewed this post and whether they had any familiarity with Brazil and their elections.

Transparency in election integrity efforts

Since 2016, Meta has said it has quadrupled the size of the global elections integrity team dealing with security and integrity efforts.⁴ According to the company, its efforts leading up to the 2022 Brazilian election included hiring experts based in Brazil and fact-checking partners who would verify Portuguese-language content. There are few details on the timeline of these efforts and whether there were sufficient resources allocated for an election of this size and magnitude.

Part of Meta's election integrity efforts includes the Elections Operations Center, which according to Meta monitors a range of issues in real time, including voter suppression content and other interference with individuals' ability to vote.⁵ Meta said it activated the Elections Operations Center for Brazil close to the October 2022 election.

When and for how long this Elections Operations Center was activated remains unclear. Although we are not certain about the extent to which special efforts were in place for the Brazil election, it is likely that in Brazil election integrity efforts were rolled back soon after the vote if what the company did after the United States election is any indication. According to media reports, Meta rolled back its election integrity efforts one month after the 2022 election in the United States.⁶

It is not entirely clear when in an election cycle it is appropriate to ramp down moderation of harmful election-related content. The aftermath of the 2020 U.S. presidential election indicated the danger of ramping down efforts to moderate election-related content in the weeks following an election, before the transition of power. The attacks on the National Congress in Brazil indicate that perhaps platforms should remain vigilant *even after* a transition of power. As elections experts write in Lawfare, both the US and Brazil “would benefit from a more prolonged period of prudence by social media companies following election day (if not permanently).”⁷ Meta should be more transparent about how it determines the conditions under which its election-related efforts, such as its “Elections Operations Center,”⁸ are activated and ramped

⁴ [Meta's work to protect the integrity of Brazil's presidential election in 2022](#)

⁵ [How Meta Is Preparing for Brazil's 2022 Elections](#)

⁶ [Facebook failed to police election misinformation among groups - The Washington Post](#)

⁷ [A Tale of Two Insurrections: Lessons for Disinformation Research From the Jan. 6 and 8 Attacks - Lawfare](#)

⁸ [How Meta Is Preparing for Brazil's 2022 Elections](#)

down. When making these determinations, it should take into account the context of the election, and the likelihood of election-related violence. Meta should also consider the possibility that a model whereby election-related operations are fully deactivated by an arbitrary date may leave the company unprepared to handle a sudden uptick in violence-inciting narratives and should continue to monitor developments once it can conclude any threat of violence has abated.

In the aftermath of the U.S. attack on the Capitol in January 2021, and the Facebook Oversight Board case regarding former President Trump, Meta rolled out its Crisis Policy Protocol to address risks to individuals during temporary high-risk incidents.⁹ Whether or not the Crisis Policy Protocol was invoked in this case and what it entails remains unclear. Meta must disclose what this Protocol entails and the criteria it uses to decide when an incident is quickly changing and requires more staff to monitor.

Transparency around enforcement of content-related policies

The questions at the center of this case are not only related to whether the company's policies are sufficient, they are also related to whether the existing suite of election and incitement of violence policies are equitably enforced. In order to facilitate accountability and oversight over the company's efforts and application of its policies, the Oversight Board and global civil society organizations require greater transparency from the company, particularly around the company's enforcement of content policies across languages and regions.

In particular, this case highlights one dimension of the general lack of transparency around how Meta's community standards are applied across the world: a dearth of documentation about how content reviewers make their decisions. At least six reviewers did not find the speech in the video to violate the company's Violence and Incitement policy, but there is no documentation as to why; reviewers are not expected to include a rationale for their content-review decisions. Moreover, we do not know why Meta ultimately reversed the findings of its moderators once the Oversight Board selected this case.

Building in some sort of mechanism by which reviewers can articulate why they made a decision, at least for posts that have been reviewed multiple times, would assist the Oversight Board and independent auditors in properly evaluating the gap between Meta's content policies and their enforcement.

Equitable enforcement of content-related policies across language and region

Part of the reason why the human reviewers did not apply the policy may have been because of their lack of familiarity with the context and language of the video. As stated earlier, narrative elements of the video had been permeating Brazilian society well before the general's remarks and the elections, which raises questions about whether the moderators reviewing the video were

⁹ [Crisis Policy Protocol | Transparency Center](#)

familiar with Brazil and the country's electoral process. An experiment conducted by Global Witness raised concerns about Meta's ability to properly detect content violating its policies against misinformation that could be classified as [voter interference](#). Global Witness submitted Brazilian Portuguese-language ads, some of which included false information about when and where to vote. All of the submitted ads were approved.¹⁰ It's not clear why the company approved all these ads but it does demonstrate a weakness in the company's application of its rules in elections and languages spoken outside of the United States context.

Experts who are fluent in local context and languages are key to ensuring equitable enforcement of community standards globally. Reviewing content is not an easy task and especially in the context of election-related speech, requires a delicate balancing of individuals' right to express themselves freely, and raises questions about government control and overreach (particularly in countries where electoral systems are captured by state actors), while also mitigating the damage of speech that suppresses voter turnout and broader civic participation.

The Santa Clara Principles, developed by global civil society organizations, may be helpful here as an example of the types of disclosures Meta can make to promote human rights and transparency: for e.g. disclosing how many reviewers work on a specific language, policy enforcement rates disaggregated by language, and when automated content analysis systems were used in flagging a post.¹¹

Meta also noted the use of artificial intelligence to identify fake accounts popping up around election period and stated that they had removed 99.7% of fake accounts before they were reported.¹² More transparency around the use of these automated systems to enforce content policies would also be welcome, for users as well as the Oversight Board. Decisions by these systems may curtail users' ability to express themselves.¹³ These tools have technical limitations that preclude them from applying content policies in a way that balances users' free expression rights, understands the motivation behind a post, and understands its likely impact. These tools are often blunt and unable to parse context, such as if a user was posting to document violence instead of promoting it. These tools are especially faulty in languages other than English where there are fewer training datasets to train these models to adequately parse different types of speech.¹⁴ As a result, deploying more automated systems in non-English languages or contexts ahead of an election without adequate investment in human oversight is not the answer due to the disparate impacts these tools can have on speakers of these languages.

¹⁰ [Facebook fails to tackle election disinformation ads ahead of tense Brazilian election | Global Witness](#)

¹¹ [Santa Clara Principles](#)

¹² [Meta's work to protect the integrity of Brazil's presidential election in 2022](#)

¹³ [No amount of "AI" in content moderation will solve filtering's prior-restraint problem - Emma J Llansó, 2020](#)

¹⁴ [Lost in Translation: Automated Content Analysis in Non-English Languages - Center for Democracy and Technology](#)

Recommendations

- Meta should create oversight mechanisms or checkpoints where human reviewers note down a reason/rationale behind a content-related decision, at least for posts that have been reviewed multiple times or have been appealed. This would enable the OB to understand the thought process behind existing policy enforcement and better identify opportunities for intervention to bring enforcement decisions closer in line with the stated policies.
- In the leadup to an election, Meta should engage a host of experts including election authorities, academics, social media platforms, journalists, election officials, and civil society, to monitor and mitigate election-related speech and extend these efforts till after the certification of the vote and in quickly changing conditions till the experts deem fit.
- Meta should disclose when, where, and in what languages they deploy automated content analysis tools to enforce election-related policies. It should also disclose to the user and the Oversight Board when a post has been taken action on by an automated system.
- Meta should disclose when, for what reason, and for how long it ramps up its election integrity efforts around an election and what that entails.
- Meta should also disclose the number of reviewers in a certain language they deploy in general, in the lead up to an election, and till when.

For more information, contact

William T. Adler, Senior Technologist, Elections & Democracy at wadler@cdt.org

Aliya Bhatia, Policy Analyst, Free Expression Project at abhatia@cdt.org