



Article

 10.1590/1809-58442025113en Open access**YOUTUBE CONTENT MODERATION:**

## Analysis of the removal of videos from the 2022 election to the January 8th coup attack

*Moderação de conteúdo no Youtube: Análise da remoção de vídeos da eleição de 2022 ao atentado golpista de 08 de janeiro**Moderación de contenidos em Youtube: Análisis de la remoción de vídeos desde las elecciones de 2022 hasta el atentado golpista del 8 de enero* Marcelo Alves dos Santos Junior*Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ – Brazil.***Editorial Details***Double-blind system***Article History:***Received: 12/23/2024**Accepted: 05/09/2025**Available online: 07/30/2025***Artigo ID:** e2025113en**Chief Editors:***Dr. Marialva Barbosa**Federal University of Rio de Janeiro (UFRJ)**Dr. Sonia Virginia Moreira**State University of Rio de Janeiro (UERJ)***Executive Editors:***Dr. Jorge C. Felz Ferreira**Federal University of Juiz de Fora (UFJF)**Dr. Ana Paula Goulart de Andrade**Federal Rural University of Rio de Janeiro (UFRRJ)***Associate Editor:***Dr. Sandro Torres de Azevedo**Federal University of Rio de Janeiro (UFRJ)***Reviewers:***Cristine Gerk**Felicity Clarke**Federal University of Rio de Janeiro (UFRJ)***Editing and XML Markup:***IR Publicações***Funding:***CNPq***How to cite:**

SANTOS JUNIOR, Marcelo Alves dos.  
*Youtube content moderation: Analysis of the removal of videos from the 2022 election to the January 8th coup attack.* São Paulo: INTERCOM - Brazilian Journal of Communication Sciences, v. 48, e2025113. <https://doi.org/10.1590/1809-58442025113en>.

**Corresponding author:**

Marcelo Alves dos Santos Junior  
[marcelo\\_alves@puc-rio.br](mailto:marcelo_alves@puc-rio.br)

**ABSTRACT:**

This study examines the content moderation policies developed and implemented by Youtube between the election period of 2022 and January 8, 2023. It builds on concepts from platform studies on the governance challenges and problems faced by these technological corporations. The study employs digital dynamic archiving methods to examine the removal status of 193,429 videos. The findings show that YouTube does not provide a clear justification for which policies were violated and why. We discuss the implications for memory and studies on contexts of risk to democracy, given that some channels have deleted more than 90% of their production. The remainder of this study evaluates the consequences for platform governance research and future research directions.

**Keywords:** Democracy, Platforms Governance, Content moderation, Digital methods, Youtube

**RESUMO**

Este texto estuda a remoção de vídeos políticos do Youtube da eleição de 2022 ao 8 de janeiro de 2023. O marco teórico se baseia nos estudos de plataformas e a literatura sobre moderação de conteúdo. Utilizamos abordagens dos métodos digitais de arquivamento dinâmico para analisar o status de remoção de uma amostra de 193.429 vídeos. Os resultados indicam que o Youtube não provê justificativas claras sobre qual política teria sido violada e por quais razões. Debates implicações para estudos sobre desinformação e integridade democrática, considerando que alguns canais deletaram mais de 90% de seus vídeos. Ao final, avaliamos as implicações para estudos sobre governança nas plataformas e próximos passos de pesquisa.

**Palavras-chave:** Democracia, Governança em plataformas, Moderação de conteúdo, Métodos digitais, Youtube

**RESUMEN**

Analizamos las políticas de moderación de contenidos desarrolladas y aplicadas por YouTube desde el periodo electoral de 2022 hasta el 8 de enero de 2023. Dialogamos con claves analíticas de estudios de plataformas sobre los retos y problemas de gobernanza llevados a cabo por estas corporaciones tecnológicas. Utilizamos aproximaciones de métodos de archivo dinámico digital para analizar el estado de eliminación de una muestra de 193.429 vídeos. Los resultados indican que Youtube no proporciona justificaciones claras sobre qué política se ha infringido y por qué motivos. Discutimos las implicaciones para la memoria y los estudios sobre contextos de riesgo para la democracia, teniendo en cuenta que algunos canales borraron más del 90% de su producción. Por último, evaluamos las implicaciones para los estudios sobre la gobernanza de las plataformas y los próximos pasos para la investigación.

**Palabras clave:** Democracia, Gobernanza de plataformas, Moderación de contenidos, Métodos digitales, Youtube

## CRediT

- Authors Contributions: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Supervision; Validation; Visualization; Writing - original draft; review and editing.
- Conflicts of Interest: The author certify that they have no commercial or associative interests that represent a conflict of interest regarding the manuscript.
- Funding: This project was funded by FAPERJ – Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro, through Program E\_13/2023 – Basic Research Support (APQ1), in accordance with the Grant and Acceptance Agreement linked to Process SEI 260003/006286/2024.

*Article submitted for similarity verification*

### **Data Availability:**

The author state that all data used in the research has been made available within the body of the article.

INTERCOM Journal encourages data sharing but, in adherence to ethical guidelines, does not require the disclosure of any means of identifying research subjects, thereby preserving their privacy. The practice of open data aims to enable the reproducibility of results and ensure unrestricted transparency in published research findings without requiring the disclosure of the subjects' identities.

*This article is published in open access (Open Access) under the Creative Commons Attribution 4.0 International (CC-BY) license. The authors retain all copyrights, transferring to Intercom: Revista Brasileira de Ciências da Comunicação the right to carry out the original publication and to keep it constantly updated.*

## Introduction

In 2022, online video-sharing platform YouTube became the main digital means of information consumption in Brazil (NEWMAN et al., 2022). Furthermore, this tool plays a key role in the construction of a far-right disinformation infrastructure (PASQUETTO et al., 2022), as its links are among the most shared on other platforms and messaging apps, such as WhatsApp and Telegram (FERNANDES NASCIMENTO et al., 2021; PIAIA; ALVES, 2020). Despite its relevance, the topic of governance on platforms is still little explored<sup>1</sup> in both Brazilian and international literature. Often unilaterally or based on court decisions, companies establish a number of acceptability parameters or algorithmic standards (FERNANDES ARAÚJO, 2021) to govern the circulation of content on their platforms. This study aims to shed light on the political and sociotechnical dimensions of governance conducted by platforms that develop and apply moderation rules to define which content should remain and which should be removed from their services.

The purpose of this article is to carry out an empirical analysis of the removal of videos by YouTube from the 2022 Brazilian general election period to the week of the attempted coup against Brazilian democratic institutions on January 8, 2023. To this end, this research is based on theoretical perspectives from platform studies in order to discuss governance and content removal models. The methodological procedures focus on approaches from digital methods, particularly the proposal for post-digital trace studies and the reconstruction of moderation scenes (VAN DIJCK; DE WINKEL; SCHÄFER, 2021; DE KEULENAAR; ROGERS, in press). We combined dynamic metadata archiving techniques for 193,429 videos through systematic queries to programmatic interfaces with data scraping from the graphical interface to record the reasons publicly stated by YouTube in case of removal.

The study seeks to answer the following questions: What are the rules of YouTube's community terms related to disinformation and democratic-electoral integrity? How did the platform moderate content in the period from the Brazilian electoral period to the insurrection of January 8, 2023? In particular, we sought to understand YouTube's role in efforts to mitigate risks to the country's electoral and democratic integrity, given the spread of disinformation, hate speech, and the crime of apology for military intervention in Brazil.

The article is organized into five parts. In the first, we briefly review the literature on governance and content moderation on digital platforms, building a typology applied in the context of YouTube. In the second section, we detail the methodological procedures, construction of the database and the technique for verifying the reasons given for the exclusion of videos. The third section shows the results of the study, indicating the changes in YouTube's electoral integrity guidelines in 2022 and the findings of the data analysis. Finally, the last section discusses the theoretical implications of the data, highlighting conclusions, limitations, and future directions for a research agenda.

## Content moderation policies on digital platforms

Digital platforms are sociotechnical infrastructures that not only host public discourses, but which also essentially reorganize them through technical arrangements, standards, and policies (GILLESPIE, 2017; VAN DIJCK; NIEBORG; POELL, 2019). In this sense, companies such as Meta and Alphabet can be understood as transnational private corporations that effectively intervene politically in the conditions of freedom of expression, modulating the boundaries of what is considered "adequate" or "acceptable" to exist and achieve visibility in the public sphere (KLONICK, 2017). These definitions and procedures are permeated by dilemmas that involve restrictions on fundamental human rights and the need to contain hate speech, illegal activities and, more recently, anti-democratic acts, disinformation and conspiracy theories (DECOOK et al., 2022).

Recent specialized literature on content moderation on digital platforms organizes practices into two categories: a) the concept of platform governance: how the systems, institutions, and state and transnational legislation regulate the possibilities of hosting content on websites; and b) the governance carried out by the platforms, i.e., the mechanisms and procedures that these companies use to curate, discipline, and sanction the activity of users and their publications (GILLESPIE, 2017). Until recently, studies on content moderation in digital media have predominantly explored the self-management activities of online forums and communities (WRIGHT; STREET, 2007). This article aims to understand the gap regarding massive commercial moderation carried out both by algorithms and self-managed interventions and by the manual work of precarious, invisible people (ROBERTS, 2019) during the 2022 electoral and post-electoral context in Brazil.

A number of content moderation policies are available on digital platforms, with different effects and

<sup>1</sup> In the Brazilian context, it is worth mentioning the Novelo Data project by Guilherme Felitti, which systematically recorded the removal of videos on YouTube during the period, indicating the removal of thousands of entries. See <https://oglobo.globo.com/blogs/sonar-a-escuta-das-redes/post/2022/11/apos-eleicoes-cana-is-bolsonaristas-retiram-do-ar-mais-de-4-mil-videos-no-youtube.ghtml>. Retrieved: January 9, 2023.

theoretical implications. Considering the degree of perceived intervention, Gillespie (2019) divides the actions into two poles of moderation: removal (hard) and filtering (soft), arguing that removal is the most extreme and noticeable intervention, while filtering (also referred to as shadow banning, i.e., unavailability of terms or accounts in searches involving certain sociodemographic or national groups) or algorithmic invisibility refers to the measures that are more difficult to detect and, often, safer from the standpoint of the image of companies. Another category of soft moderation is content labeling, i.e., the inclusion of a superficial text or note objecting the information or signaling that it is disputed by other sources (CRAWFORD; GILLESPIE, 2016).

Goldman (2021) proposes a typology of content moderation with five categories: a) content regulation: interventions at the level of the published message, including removal, suspension, relocation, warning, or captioning; b) account regulation: blocking users, suspension of the posting feature and removal of credibility marks, such as the blue checkmark; c) reduction of visibility: algorithmic control of the circulation of publications and sanctions applied as a result of fact-checking, exclusion of search responses (shadow banning); d) demonetization of accounts and imposition of financial fines; and e) others, such as YouTube's system of applying cumulative sanctions, which implies suspension of the account.

The theoretical typologies proposed by Gillespie (2019), Goldman (2021) and Grimmelman (2015) suggest how governance methods on platforms are broad and varied. YouTube is an example of this multifaceted approach, with several moderation mechanisms. In 2019, the company published the document "The Four Rs of Responsibility," detailing the governance procedures for content considered harmful. The platform stipulates four dimensions of measures: remove, raise, reward and reduce. According to the platform, the creation of moderation rules is relative to each context, explaining that, in more problematic cases, the development and review of policies may take several months, involving consultation with experts outside of YouTube, content creators, and regional specificities (YOUTUBE, 2019a). A comparative analysis of the terms of service of mainstream and alternative platforms indicated that YouTube has clearer, more intelligible, granular and illustrative rules than other platforms (SINGHAL et al., 2022).

Removal occurs at three levels: a) channels: the account and all its videos are removed at once; b) videos: analysis carried out based on the content or title and description metadata; and c) comments written by users. Removal is presented by YouTube as the main sanction against content that violates the platform's community guidelines, as the speed of removal and the reduction in exposure time of videos that fail to comply with the rules are the central metrics of the company's transparency report.

Monetization, in turn, is introduced as a positive measure to reward the creation of content classified by the company as reliable and authoritative. In this way, the power of platforms that act as infrastructures that classify and organize information, stakeholders, and behaviors becomes clear, drawing symbolic and concrete lines between what is acceptable and unacceptable (BOWKER; STAR, 2000). The exercise of classification becomes a sanction when YouTube revokes these privileges and demonetizes a channel due to violation of community rules (CAPLAN; GILLESPIE, 2020). Conversely, this reward can become a mechanism to encourage the spread of misinformation and hate speech in cases of inconsistent application of the policy or incorrect or biased judgments, as was the case of Jovem Pan, which, despite being a major exponent of communication on right-wing networks, often producing misinformation and hate speech (SANTINI et al., 2023), received millions of reais on YouTube through the Google News Initiative program and was only demonetized in late November 2022.

The last two measures represent the main soft moderation actions: up-ranking and demotion/down-ranking. Up-ranking is the effect of highlighting the search engine rankings and the video recommendation algorithm. This effect of algorithmic expansion of visibility is applied to channels that are classified by YouTube as credible sources on the topics of "news, science, and historical events" (YOUTUBE, 2019b, n.p.). There are two major epistemic processes that should be noted. The first is the understanding that not all topics are subject to amplification, with the implication of a selection between categories that are privileged to the detriment of others; and the second is the platform's power to define which type of source is categorized as having "credibility." Possible problems in this classification can also transform a reward into an algorithmic incentive for the visibility of misinformation, particularly considering that Jovem Pan and Fox News were listed by YouTube itself as trusted sources: "In 2017, we began prioritizing authoritative voices, including news sources such as CNN, Fox News, Jovem Pan, India Today, and The Guardian, for news and information searches in search results and in 'Watch Next' panels" (YOUTUBE, 2019, n/p).

Finally, visibility reduction (down-ranking) is a sanction used for content that is classified as borderline, i.e., it is perceived as problematic by the platform but does not violate specific terms of YouTube's content moderation policy. The company cites harmful misinformation, which is classified with the aid of external fact-checkers and

experts (YOUTUBE, 2019b). According to YouTube, consensus decisions are now used as a database to train machine learning algorithms that will automate the identification of borderline videos and reduce their potential reach.

## Methodological procedures

Applied studies on content moderation on digital platforms face two challenges: the lack of public transparency and the difficulty of accessing data. Information on the application of governance measures is not available in the programmatic interfaces of digital platforms. Therefore, researchers have recently developed methods to digitally reconstruct the removal metadata. In particular, De Keulenaar and Rogers (in press) argue for a shift in the theoretical perspective of digital methods on the analysis of traces on digital platforms. According to the authors, research on the digital often treats data as information that is “raw” and “unobstructed” by public opinion. On the contrary, De Keulenaar and Rogers emphasize the artificiality of digital data in several dimensions, particularly in relation to governance measures that shed light on the sociotechnical effects of the infrastructural agency of platforms. In this sense, it is a matter of analyzing, as a priority, the effects of the platform in the modulation of phenomena and messages as an object and locus of research.

To this end, the authors propose the development of methods for systematic and continuous “dynamic archiving” of digital platforms in order to “reconstruct the scenes” of moments before and after the application of content moderation measures. Conversely, this involves using mechanisms such as the Wayback Machine tool to trace the transformations in YouTube’s moderation guidelines; and, on the other, conduct reverse engineering on the moderation traces through the combination of data extraction techniques. This approach emphasizes study opportunities in a post-API<sup>2</sup> context (BRUNS, 2019) and restrictions on data transfer policies by platforms (D’ANDRÉA, 2021), as it builds databases that are not otherwise available.

## List of channels

Related works compose samples using query terms in order to archive publications on a specific topic or hashtag, such as COVID-19 or *#stopthesteal* (DE KEULENAAR; BURTON; KISJES, 2021; SUZOR, 2020). Nevertheless, the problem with this research design lies in the fact that moderation obfuscation tactics, particularly lexical neologisms (DE KEULENAAR and ROGERS, in press), significantly complicate the task of identifying videos that aim to evade detection and which are, therefore, more difficult to capture in the collection, such as those that use terms such as “*vachina*,” “*v4cina*,” “*v4c1na*,” and derivatives, all derived from Portuguese “*vacina*” (vaccine). Following the approach of Rauchfleisch and Kaiser (2021), we sampled 849 YouTube channels that posted videos on political-electoral topics in 2022. The list is the result of a snowball sampling process carried out by the Channel Network module of the YouTube Data Tools tool, which analyzes how channels follow and recommend each other. We carried out two rounds of network expansion at depth level 1 with manual verification control based on two criteria: a) Brazilian Portuguese language; and b) have published at least one video about the Brazilian political-electoral context in the year.

## Database composition by dynamic archiving

Next, we created a cloud-hosted system with a script written by the author using the tubeR package of the R programming language to access the YouTube API<sup>ube</sup> and collect the last 10 to 40 videos from the channel list every six hours, in the period between July 15, 2022 and January 15, 2023. The metadata was consolidated by eliminating duplicates, resulting in a database of 193,429 videos. In the first week of February 2023, we performed new queries to the YouTube API to update the metrics of all publications. Through the programmatic interface, videos without statistics are unavailable on the site (DE KEULENAAR; BURTON; KISJES, 2021; SUZOR, 2020). We created a second script to control a browser in an anonymous session<sup>3</sup> with cookie deletion in each session to access all the URLs of unavailable videos and scrape the text of the reasons reported by YouTube.

Considering that there are different reasons for deleting a video, many of which are very infrequent (SUZOR, 2020), we created the following categories to group the removal messages:

2 A term used to contextualize a time when public Application Programming Interfaces (APIs), traditional data sources for researchers, are closing, as is the case with social media platforms Instagram, Facebook, and Twitter.

3 We followed the best practices from specialized literature for autonomous and systematic browser testing. The automation was performed by the Selenium WebDriver tool (<https://www.selenium.dev/>) and operated in a Docker virtual environment (<https://www.docker.com/>) for stability and replicability purposes.

- Moderated by the author: messages warning that the channels themselves made the videos private or deleted the content;
- Violation: messages notifying that the video violated a community rule or YouTube terms of service;
- Account closed: messages regarding the deplatforming of the channel as a whole and deletion of the videos as a result;
- Copyright: violation of the copyright policy; and
- Undefined:<sup>4</sup> default message stating “this video is unavailable” without further details.

### **Content moderation policies**

We analyzed the changes relevant to the Brazilian election between 2022 and 2023 made to the text on the YouTube Elections Misinformation Policies page and supplemented the information with data from the report “*YouTube e as Eleições Brasileiras de 2022: Retrospectiva*” (“YouTube and the 2022 Brazilian Elections: A retrospective”) prepared by the platform to provide transparency on moderation efforts within the scope of cooperation with Brazil’s Superior Electoral Court (*Tribunal Superior Eleitoral – TSE*) to combat disinformation.

### **Results**

In this section, we present the results of the analyses carried out, organized into four headings: 1) Content moderation policies in Brazil; 2) Reasons expressed by YouTube; 3) Time series of removal; and 4) Moderation proportions by channel.

### **Changes in YouTube’s electoral integrity policies**

The first aspect to highlight is the fact that YouTube did not have a policy ready for the Brazilian context in 2022 and was in the process of creating or changing the rules during the first half of the year or even during the electoral campaign. Content moderation policies on the election focus on three categories: a) policies on deceptive practices; b) violence, hate speech, and harassment; and c) impersonation and false interaction. Specifically, the first rule prohibits the use of manipulated content, voter suppression, false information about candidate eligibility, incitement to interfere in democratic processes (queues, breach of electronic security), and claims that the elections have been rigged or that flaws altered the election results.

The rules, however, are only applied after institutional certification of the election results. On April 21, 2022, the policy was updated to add the electoral integrity clause against claims of fraud in the 2018 elections. On August 10, this same rule was also applied to the 2014 election. In turn, only on October 31, the day after the results of the second round were announced by the TSE, YouTube expressly included the 2022 election in the electoral integrity policy.

The platform has a rule for cumulative penalties in which channels that receive three strikes for content removal for violating the terms of service within 90 days are to be permanently banned. Nevertheless, the changes to the electoral integrity policies were followed by 30-day grace periods in which videos that violated the terms could be removed from the site, but would not count towards permanent suspension. In fact, since there were three changes on different dates, no cumulative sanctions were applied for ninety days, including the month after the election. “This means that content that violated our guidelines posted between October 31, 2022 and November 30, 2022 was removed from our platform, but the channels did not receive any penalty other than the removal of these videos” (YOUTUBE, 2023, p.3).

It is worth questioning why the corporation made different decisions for the 2014 and 2018 elections, which doubled its grace period to 60 days in a pivotal year for Brazilian democracy. Moreover, it is noteworthy that the recognition of the fairness of the 2022 election was only included in the policy after the second round, which allowed a period of non-applicability of the policy throughout the year, since, until the promulgation of the result, the terms only referred to 2014 and 2018.

YouTube indicated in its transparency report that over 10,000 videos and 2,500 channels were removed for having violated the policies on the elections in Brazil. According to the company’s data, more than 84% were deleted before reaching 100 views. Nevertheless, beyond dynamic archiving efforts such as the one carried out in this

<sup>4</sup> We considered the “Undefined” category as an exclusion made by the platform without specifying the reason. We understand that it is not an action taken by the channel itself, which is covered in other messages. Nevertheless, we did not find public documentation clearly detailing how to interpret these cases and thus followed related studies. This is a major challenge, also criticized by the specialized literature.

research, there is no way to obtain this database and carry out an independent verification or assessment of the report.

### **Reasons expressed by YouTube**

We found 29,755 unavailable videos, accounting for 15.4% of a sample of 193,429, totaling 870.4 million views and an average of 29,200 per analysis unit. Considering the URLs that were offline, we present the reasons given by YouTube on the page (TABLE 1).

**TABLE 1** - Video status, reason, and views

Status	Frequency	Views (total)	Views (average)
Online	163,690 (84.6%)	5,644,636,495	34,483,7
Undefined	17,384 (9.0%)	485,843,722	27,947,75
Channel	11,769 (6.1%)	371,879,172	31,598,2
Account terminated	421 (0.2%)	5,796,027	13,767,29
Violation	142 (0.1%)	4,395,405	30,953,56
Copyright	22 (0.0%)	2,474,471	112,476

**SOURCE** – Author’s own work based on data extracted from YouTube.

The results show that 9% of the removals only presented the message “*This video isn’t available anymore,*” without any further details or justifications; another 6.1% were related to the actions of the channels themselves, considering that 7,709 videos were made private and 4,060 were deleted. These values are important and indicate that removal was not a consequence of the platforms’ actions, but rather a preventive act by the accounts themselves.

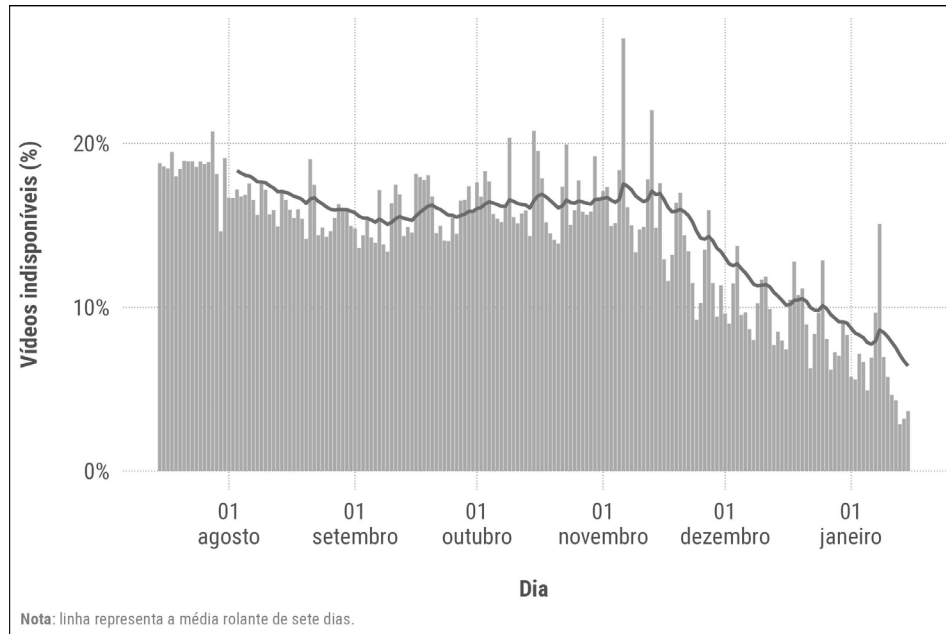
It should be noted that only 142 disabled videos contain a justification for violating a YouTube policy. The lack of transparency remains, however, as 118 cases have a generic message, such as: “*This video has been removed for violating YouTube’s Community Guidelines,*” while another 15 say “*This video has been removed for violating YouTube’s Terms of Service.*” The two most frequent justifications that expressly indicate which rule was broken are: “*This video has been removed for violating YouTube’s policy on harassment and bullying,*” presented five times, and “*This video has been removed for violating YouTube’s policy on violent or graphic content,*” presented four times. It is noteworthy that no video has a message explaining the exclusion due to a violation of the electoral integrity policy or hate speech, which prevents an in-depth analysis and review of the platform’s moderation criteria.

Considering the subsample that has an explicit violation message, the video with the greatest reach was “ACABOU de ACONTECER 30.10 #rio” from the channel #MANOTOKIO2020, removed for violating the violent or explicit content guideline, which, in our last record, had reached 1.1 million views. It should be noted that even videos with an electoral topic identified in the title did not present information compatible with the removal, such as: “*FA x TSE – Teremos eleições [sic] confiáveis?*” (“Armed Forces vs. Superior Electoral Tribunal – Will we have fair elections?”) posted by Brasil pela Direita, “*BRAZIL WAS STOLEN – Auditoria Privada das Eleições 2022*” (“Brazil Was Stolen – Private Audit of the 2022 Election” by Tramonte, or “*Auditores argentinos atestam fraudes nas eleições brasileiras*” (“Argentine auditors attest fraud in the Brazilian elections”) by Carlos Ferrari.

### **Temporal dynamics of removal**

Another aspect to be analyzed is the date on which the videos that are unavailable were posted. In theory, this could have occurred when episodes of violations to guidelines or changes in the investment of resources for moderation by the platform are concentrated. FIGURE 1 shows the daily percentage of unavailable publications at the time of verification as of February 2023.

**FIGURE 1:** Percentage of unavailable videos per day and seven-day rolling average line.

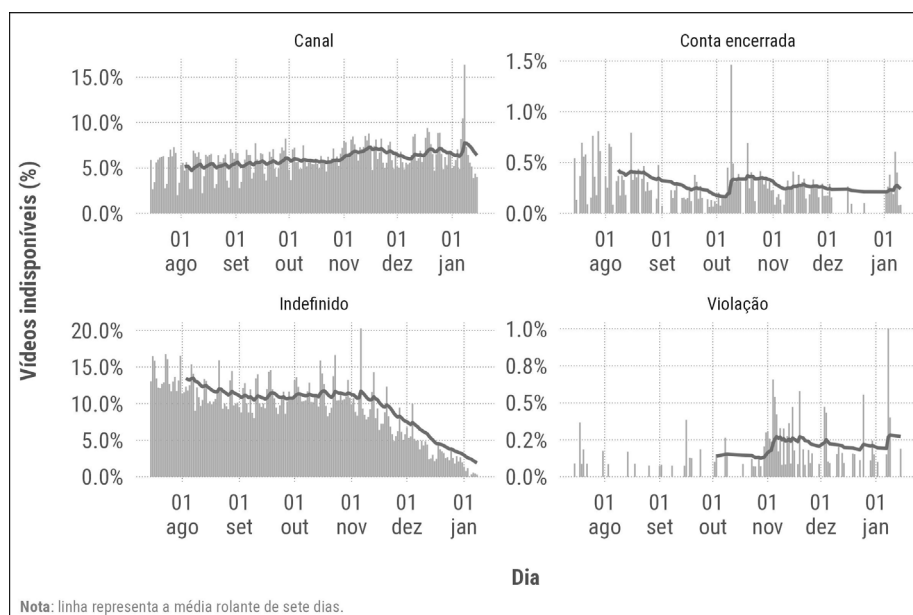


**SOURCE** – Author’s own work based on data extracted from YouTube. Note: Line represents seven-day rolling average.

The results of the time series indicate that July was the month with the highest proportion of videos removed (18.66%). There was stability in the months of August (16%), September (15.5%), and October (16.6%). Despite being the month of the first and second round of the election, a time when, hypothetically, messages would be more extreme, there was no significant increase in removals in October. There were two peaks: November 6, with 29.1% of the offline sample, and November 13, with 23.1%. In total, on ten dates, over 20% of the videos were unavailable.

The main finding of the time series analysis is the continuous and significant drop that occurs after the second round of the election. Only 6.24% of the content was offline in January. We can raise three competing hypotheses to interpret this data: 1) With the election results, channels reduced postings of political-electoral content or learned to obfuscate messages that directly questioned Lula’s victory and the fairness of the electoral process; 2) although the platform tries to remove violations quickly, this procedure takes a long time; or 3) YouTube reduced the dedication of resources, staff, and content moderation systems that were activated to operate during the months of the campaign. FIGURE 2 disaggregates the reasons for the unavailability of the videos to shed light on these issues.

**FIGURE 2:** Total number of unavailable videos per day considering the reason for removal.



**SOURCE** – Author’s own work based on data extracted from YouTube. Note: Line represents seven-day rolling average.



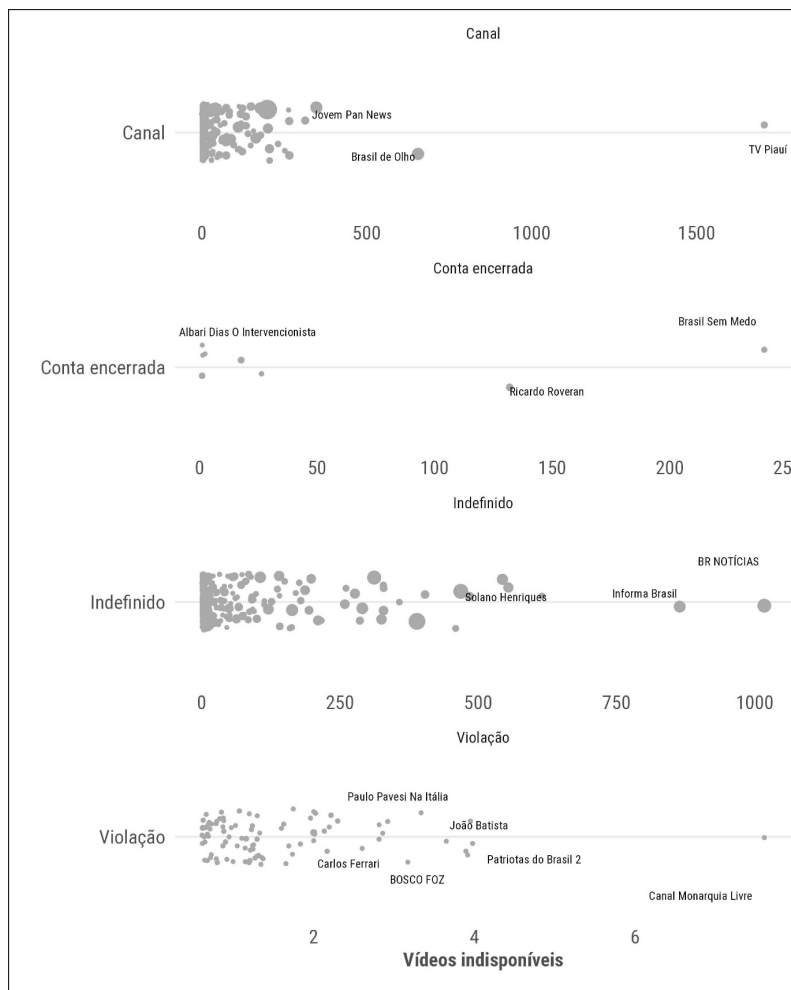
When the reasons for removals are observed separately, a reversal of trends can be seen. Throughout the election period, the volume of content removed by YouTube is higher. This pattern begins to change from the day after the second round of voting. In fact, from November 1 onwards, there is a constant drop in the “Undefined” label and a maintenance with a slight decrease in the measures taken by channel management. From the second week of December onwards, the lines reverse, reaching a greater discrepancy on January 8, the day of the vandalism of the buildings of Brazil’s Three Powers in Brasília, when 163 videos were deleted by those who posted them, while only 10 were removed by YouTube for violating the rules, four without specific reasons, and three due to account closure. This indicates that the channels continued to post messages that potentially challenged YouTube’s rules and took measures to manage the risk on an a posteriori basis, either deleting the content or making it as private, while the platform relaxed its moderation measures.

**Video removal by channels**

Some channels were completely removed, which made it impossible to access all the videos produced. Considering the sample of this research, 49 accounts (5.7%) had 100% of their content unavailable at the time of verification. This means that if a study were to start collecting data from February 2023 onwards, at least 49 channels would have been removed from the YouTube database, becoming inaccessible for subsequent research. More generally, 116 channels (15%) have more than 75% of their videos offline.

When looking at the details of the data, it is clear that a number of channels deleted almost all of the publications made during the period of this research. The TV Piauí channel, for example, deleted or made private 1,705 videos (99.88%) of the sample collected. The same pattern is repeated for Patriota em Ação (99.12%), Pastora Valdirene Moreira (99.1%), Renato R Gomes (98.33%), TV Bolsonaro Rumo à 2022 (98.05%), and Brasil de Olho (97.47%).

**FIGURE 3:** Total number of videos removed from YouTube grouped by channel



SOURCE – Author’s own work based on data extracted from YouTube.



Removals due to violations are less common. The account with the most exclusions in this category was Canal Monarquia Livre, with eight violations (0.51%). In total, 13 channels reached the benchmark of three violations in the period studied by this article. Nevertheless, it is not possible to assess more clearly the policy of applying strikes that imply permanent blocking of the channel based on the grace periods and the 90-day deadline.

## Conclusions

This article analyzed the application of content moderation measures on YouTube between July 15, 2022 and January 15, 2023. Based on theoretical and methodological debates that re-propose the study of digital traces to elucidate the effects and agencies of platforms (DE KEULENAAR and ROGERS, in press), the research design reconstructs the scenes of removed videos through dynamic archiving of metadata and scraping of the reasons for the unavailability of content. The research aims to contribute to the theoretical framework on digital platform governance, particularly regarding content moderation and political-electoral disinformation (GILLESPIE, 2017; GOLDMAN, 2021).

The specialized bibliography on governance on platforms carries out a critical discussion of the exercise of private power by these technology corporations as active editors of discourses in the public sphere, without democratically constituted procedures, public legitimacy, or adaptation to national specificities (SCHARLACH; HALLINAN; SHIFMAN, 2023). Although YouTube's moderation policies are among the most intelligible (SINGHAL, et al, 2021), it should be noted that most videos are removed without publicly explaining the reasons and policies that were broken, which reduces the transparency and accountability of the governance framework. The measures implemented by the platforms (GILLESPIE, 2019) must be complemented by legislation debated with civil society and approved by the National Congress, in order to regulate content removal practices (KLONICK, 2017). Additionally, in line with studies on the temporal dimension of governance (SUZOR, 2020), we emphasize the relevance of the post-electoral period in contentious elections, demonstrating that YouTube reduced content moderation efforts during the period of intense contestation of the election results and mobilizations before military facilities.

The results highlight that it is extremely challenging to research contexts of political-electoral tension with great risks to democratic integrity on the platform. There is a problem of recovering digital memory and accessing data for conducting research into anti-democratic movements, disinformation, and political violence, as almost one third of the videos ceased to exist on YouTube, just one month after the phenomenon (RAUCHFLEISCH; KAISER, 2021). Furthermore, the findings reinforce the need to approve regulations and procedural mechanisms to increase public transparency regarding the moderation decisions made by the platforms and subsequent preservation of the database for academia (HARTMANN et al., 2023; SUZOR, 2020).

Platform governance is a relevant research agenda for media, technology, law, and social science studies. Future studies could promote a more systematic analysis of the content of removed videos compared to those that are still online. Part of the analysis should question the consistency of the application of the platform's moderation policies. Nevertheless, it is essential to problematize these regulations in light of normative concepts of democratic theory or the conceptual framework of human rights, in order to advance in the proposition of governance models and community standards that are more participatory and representative of substantive values and national contexts.

## References

- ALVES, Marcelo. Clones do YouTube: replataformização da irrealidade e infraestruturas de desinformação sobre a Covid-19. **Revista Fronteiras**, v. 23, n. 2, 2021.
- BOWKER, G. C.; STAR, S. L. **Sorting things out: Classification and its consequences**. [s.l.] MIT press, 2000.
- BRUNS, A. After the 'APIcalypse': social media platforms and their fight against critical scholarly research. **Information, Communication & Society**, v. 22, n. 11, p. 1544–1566, 19 set. 2019.
- CAPLAN, R.; GILLESPIE, T. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. **Social Media + Society**, v. 6, n. 2, p. 205630512093663, abr. 2020.
- D'ANDRÉA, C. Para além dos dados coletados: políticas das APIs nas plataformas de mídias digitais. **MATRIZES**, v. 15, n. 1, p. 103–122, 8 jun. 2021.
- DE ALMEIDA FONSECA, G.; D'ANDRÉA, C. Governança e mediações algorítmicas da plataforma YouTube durante a pandemia de COVID-19. **Dispositiva**, v. 9, n. 16, p. 6–26, 2020.

- DE KEULENAAR, E.; BURTON, A. G.; KISJES, I. Deplatforming, demotion and folk theories of Big Tech persecution. **Fronteiras - estudos midiáticos**, v. 23, n. 2, p. 118–139, 14 set. 2021.
- DE KEULENAAR, E.; ROGERS, R. **The Return of Trace Research for the Study of Platform Effects**. no prelo.
- DECOOK, J. R. et al. Safe from “harm”: The governance of violence by platforms. **Policy & Internet**, v. 14, n. 1, p. 63–78, 2022.
- FERNANDES ARAÚJO, W. Norma algorítmica como técnica de governo em Plataformas Digitais: um estudo da Escola de Criadores de Conteúdo do YouTube. **Revista Fronteiras**, v. 23, n. 1, 2021.
- FERNANDES NASCIMENTO, L. et al. Poder oracular e ecossistemas digitais de comunicação: a produção de zonas de ignorância durante a pandemia de Covid-19 no Brasil. **Revista Fronteiras**, v. 23, n. 2, 2021.
- GILLESPIE, T. The politics of ‘platforms’. **New Media & Society**, v. 12, n. 3, p. 347–364, maio 2010.
- \_\_\_\_\_. Governance of and by platforms. **The SAGE handbook of social media**, p. 254–278, 2017.
- \_\_\_\_\_. **Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media**. [s.l.] Yale University Press, 2019.
- \_\_\_\_\_. et al. Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. **Internet Policy Review**, v. 9, n. 4, 21 out. 2020.
- GOLDMAN, E. Content moderation remedies. **Mich. Tech. L. Rev.**, v. 28, p. 1, 2021.
- GRIMMELMANN, J. The virtues of moderation. **Yale JL & Tech.**, v. 17, p. 42, 2015.
- HARTMANN, I. et al. **Moderação de conteúdo online: contexto, cenário brasileiro e suas perspectivas regulatórias**. Rio de Janeiro: Alameda, 2023.
- KLONICK, K. The new governors: The people, rules, and processes governing online speech. **Harv. L. Rev.**, v. 131, p. 1598, 2017.
- NEWMAN, N. et al. Reuters Institute Digital News Report 2022.
- PASQUETTO, I. V. et al. Disinformation as Infrastructure: Making and maintaining the QAnon conspiracy on Italian digital media. **Proceedings of the ACM on Human-Computer Interaction**, v. 6, n. CSCW1, p. 1–31, 2022.
- PIAIA, V.; ALVES, M. Abrindo a caixa preta: análise exploratória da rede bolsonarista no WhatsApp. **Intercom: Revista Brasileira de Ciências da Comunicação**, v. 43, p. 135–154, 2020.
- RAUCHFLEISCH, A.; KAISER, J. Deplatforming the Far-right: An Analysis of YouTube and BitChute. **SSRN Electronic Journal**, 2021.
- ROBERTS, S. T. **Behind the screen: content moderation in the shadows of social media**. New Haven: Yale University Press, 2019.
- ROGERS, R. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. **European Journal of Communication**, v. 35, n. 3, p. 213–229, 2020.
- SANTINI, Rose Marie; SALLES, Débora; MATTOS, Bruno. Recommending instead of taking down: YouTube hyperpartisan content promotion amid the Brazilian general elections. **Policy & Internet**, 2023.
- SCHARLACH, R.; HALLINAN, B.; SHIFMAN, L. Governing principles: Articulating values in social media platform policies. **New Media & Society**, p. 14614448231156580, 11 nov. 2023.
- SINGHAL, M. et al. **SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice**. arXiv, 27 out. 2022. Disponível em: <<http://arxiv.org/abs/2206.14855>>. Acesso em: 11 nov. 2023
- SUZOR, N. Understanding content moderation systems: new methods to understand internet governance at scale, over time, and across platforms. Em: WHALEN, R. (Ed.). **Computational Legal Studies**. [s.l.] Edward Elgar Publishing, 2020. p. 166–189.
- VAN DIJCK, J.; DE WINKEL, T.; SCHÄFER, M. T. Deplatformization and the governance of the platform ecosystem. **New Media & Society**, p. 14614448211045662, 2021.
- WRIGHT, S.; STREET, J. Democracy, deliberation and design: the case of online discussion forums. **New media & society**, v. 9, n. 5, p. 849–869, 2007.
- YOUTUBE. **The Four Rs of Responsibility, Part 1: Removing harmful content**. Disponível em: <<https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-remove/>>. Acesso em: 5 mar. 2023a.

**YOUTUBE. The Four Rs of Responsibility, Part 2: Raising authoritative content and reducing borderline content and harmful misinformation.** Disponível em: <<https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-raise-and-reduce/>>. Acesso em: 5 mar. 2023b.