


Artigo

 10.1590/1809-58442025113pt Open access

## MODERAÇÃO DE CONTEÚDO NO YOUTUBE:

### Análise da remoção de vídeos da eleição de 2022 ao atentado golpista de 08 de janeiro

*Youtube content moderation: Analysis of the removal of videos from the 2022 election to the January 8th coup attack*

*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

*Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ - Brasil.*

#### Detalhes Editoriais

Sistema duplo cego

#### Histórico do Artigo:

Recebido: 23/12/2024

Aceito: 09/05/2025

Disponível online: 30/07/2025

Artigo ID: e2025113pt

#### Editoras Chefes:

Dra. Marialva Barbosa  
Universidade Federal do Rio de Janeiro,  
UFRJ

Dra. Sonia Virginia Moreira  
Universidade Estadual do Rio de Janeiro,  
UERJ

#### Editores Executivos:

Dr. Jorge C. Felz Ferreira  
Universidade Federal de Juiz de Fora, UFJF

Dra. Ana Paula Goulart de Andrade  
Univ. Federal Rural do Rio de Janeiro, UFRJ

#### Editor Associado:

Dr. Sandro Torres de Azevedo  
Universidade Federal do Rio de Janeiro,  
UFRJ

#### Revisoras:

Cristine Gerk (português)  
Felicity Clarke (Inglês)  
Universidade Federal do Rio de Janeiro,  
UFRJ

#### Editoração e marcação XML:

IR Publicações

#### Financiamento:

CNPq

#### Como citar:

SANTOS JUNIOR, Marcelo Alves dos. Análise da remoção de vídeos da eleição de 2022 ao atentado golpista de 08 de janeiro. São Paulo: INTERCOM - Revista Brasileira de Ciências da Comunicação, v. 48, e2025113. <https://doi.org/10.1590/1809-58442025113pt>.

#### Autor de contato:

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

#### 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 implicam 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.

#### 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.

#### 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

- Conflitos de Interesse: o autor certifica que não tem interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- Contribuições dos autores: Administração do projeto; Análise formal; Conceitualização; Curadoria de dados; Escrita - rascunho original; Investigação; Metodologia; Supervisão; Validação; Visualização; revisão e edição.
- Financiamento: Este projeto foi financiado pela FAPERJ – Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, por meio do Programa E\_13/2023 - Auxílio Básico à Pesquisa (APQ1), conforme Termo de Outorga e Aceitação de Auxílio vinculado ao Processo SEI 260003/006286/2024.

*Artigo submetido à verificação de similaridade*

### **Disponibilidade dos Dados:**

todos os dados utilizados na pesquisa foram disponibilizados no corpo do artigo.

A REVISTA INTERCOM incentiva o compartilhamento de dados mas, por observância a ditames éticos, não demanda a divulgação de qualquer meio de identificação de sujeitos de pesquisa, preservando a privacidade dos sujeitos de pesquisa. A prática de open data é viabilizar a reproducibilidade de resultados, e assegurar a irrestrita transparência dos resultados da pesquisa publicada, sem que seja demandada a identidade de sujeitos de pesquisa.

*Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC- BY). Os autores retêm todos os direitos autorais, transferindo para a Intercom: Revista Brasileira de Ciências da Comunicação o direito de realizar a publicação original e mantê-la sempre atualizada.*

## Introdução

Em 2022, o Youtube se tornou o principal meio digital de consumo de informação no Brasil (NEWMAN et al., 2022). Além disso, cumpre um papel central na construção de uma infraestrutura desinformativa (PASQUETTO et al., 2022) de extrema-direita, na medida que seus links estão entre os mais compartilhados em outras plataformas e em aplicativos de troca de mensagens, como WhatsApp e Telegram (FERNANDES NASCIMENTO et al., 2021; PIAIA; ALVES, 2020). Apesar de tamanha relevância, a temática da governança nas plataformas ainda é pouco explorada<sup>1</sup> pela bibliografia nacional e internacional. De forma muitas vezes unilateral ou com base em decisões judiciais, empresas estabelecem diversos parâmetros de aceitabilidade ou normas algorítmicas (FERNANDES ARAÚJO, 2021) para governar a circulação de conteúdo em seus sites. Esta pesquisa visa jogar luz sobre a dimensão política e sociotécnica da governança realizada pelas plataformas que desenvolvem e aplicam regras de moderação para definir quais conteúdos devem permanecer e quais devem ser excluídos de seus serviços.

O objetivo deste artigo é analisar empiricamente a remoção de vídeos pelo Youtube durante o período eleitoral de 2022 até a semana da tentativa de golpe contra as instituições democráticas brasileiras em 8 de janeiro de 2023. Para isso, essa pesquisa se fundamenta em perspectivas teóricas dos estudos de plataforma a fim de debater os modelos de governança e de remoção de conteúdo. Já os procedimentos metodológicos se debruçam em abordagens dos métodos digitais, particularmente, a proposta de estudos pós-traços digitais e de reconstituição de cenas de moderação (VAN DIJCK; DE WINKEL; SCHÄFER, 2021; DE KEULENAAR; ROGERS, no prelo). Combinamos técnicas de arquivamento dinâmico de metadados de 193.429 vídeos por meio de consultas sistemáticas às interfaces programáticas com raspagem de dados da interface gráfica para registrar as razões publicamente expostas pelo Youtube em caso de remoção.

A pesquisa busca responder às seguintes questões: quais são as regras dos termos de comunidade do Youtube relacionadas à desinformação e à integridade democrático-eleitoral? Como a plataforma realizou a moderação de conteúdo do período eleitoral brasileiro até a insurreição golpista de 08 de janeiro de 2023? Particularmente, buscamos entender o papel do Youtube nos esforços de mitigação dos riscos à integridade eleitoral e democrática do país, tendo em vista o espalhamento de desinformação, discurso de ódio e crime de apologia à intervenção militar no Brasil.

O artigo está organizado em cinco partes. Na primeira, revisamos brevemente a literatura sobre governança e moderação de conteúdo em plataformas digitais, construindo uma tipologia aplicada no contexto do Youtube. Na segunda seção, detalhamos os procedimentos metodológicos, construção do banco de dados e a técnica de verificação das razões expostas para a exclusão de vídeos. A terceira demonstra os resultados da pesquisa, indicando as transformações nas diretrizes de integridade eleitoral do Youtube em 2022 e os achados da análise de dados. Por fim, o último trecho debate as implicações teóricas dos dados, aponta conclusões, limitações e caminhos futuros para uma agenda de investigação.

## As políticas de moderação de conteúdo em plataformas digitais

Plataformas digitais são infraestruturas sociotécnicas que não somente hospedam discursos públicos, mas, essencialmente, os reorganizam por meio de agenciamentos técnicos, normas e políticas (GILLESPIE, 2017; VAN DIJCK; NIEBORG; POELL, 2019). Nesse sentido, empresas como Meta e Alphabet podem ser entendidas como corporações privadas transnacionais que, efetivamente, intervêm politicamente nas condições de liberdade de expressão, modulando as fronteiras do que é considerado “adequado” ou “aceitável” para existir e atingir visibilidade na esfera pública (KLONICK, 2017). Essas definições e procedimentos são atravessados por dilemas que se tensionam restrições de direitos humanos fundamentais e a necessidade de conter discurso de ódio, atividades ilegais e, mais recentemente, atos antidemocráticos, desinformação e teorias da conspiração (DECOOK et al., 2022).

A literatura especializada recente sobre moderação de conteúdo em plataformas digitais organiza as práticas em duas categorias: a) o conceito de governança das plataformas: como os sistemas, instituições e legislações estatais e transnacionais que regulam as possibilidades de hospedagem de conteúdo nos sites; e b) a governança realizada pelas plataformas, isto é, mecanismos e procedimentos que estas empresas acionam para curar, disciplinar e sancionar a atividade de usuários e suas publicações (GILLESPIE, 2017). Até então, estudos sobre moderação de conteúdo em meios digitais predominantemente exploram as atividades de auto-gestão de fóruns e comunidades (WRIGHT; STREET, 2007). Este artigo se propõe a compreender a lacuna acerca da moderação comercial massiva realizada tanto por algoritmos e intervenções automáticas, quanto pelo trabalho manual de pessoas precarizadas e

<sup>1</sup> No contexto nacional, vale menção ao projeto da *Novelo Data* de Guilherme Felitti que registrou sistematicamente a remoção de vídeos no Youtube durante o período, apontando o apagamento de milhares de entradas. Ver <https://oglobo.globo.com/blogs/sonar-a-escuta-das-redes/post/2022/11/apos-eleicoes-canais-bolsonaristas-retiram-do-ar-mais-de-4-mil-videos-no-youtube.ghtml>. Acesso em 09 de jan. 2023.

invisibilizadas (ROBERTS, 2019) durante o contexto eleitoral e pós-eleitoral de 2022 no Brasil.

Existem diversas políticas de moderação de conteúdo em plataformas digitais com efeitos e implicações teóricas distintas entre si. Levando em consideração o grau de intervenção percebida, Gillespie (2019) divide as ações em dois polos de moderação: remoção (*hard*) e filtragem (*soft*), argumentando que a exclusão é a intervenção mais extrema e perceptível, enquanto que a filtragem (também chamada de *shadow banning*, isto é, a indisponibilidade de termos ou contas na busca em determinados recortes sociodemográficos ou nacionais) ou invisibilidade algorítmica refere-se às medidas mais dificilmente detectáveis e, muitas vezes, mais seguras do ponto de vista da imagem das empresas. Outra categoria de moderação leve é a rotulagem de conteúdos, ou seja, a adição de um texto ou marca superficial de objeção ou de sinalização de que a informação é disputada por outras fontes (CRAWFORD; GILLESPIE, 2016).

Já Goldman (2021) propõe uma tipologia da moderação de conteúdo com cinco categorias: a) regulação de conteúdo: intervenções no nível da mensagem publicada, incluindo remoção, suspensão, relocação, aviso ou legenda; b) regulação de contas: bloqueio de usuários, suspensão da função de postar e remoção de marcas de credibilidade, como o símbolo azul; c) redução de visibilidade: controle algorítmico da circulação das publicações e sanção aplicada em consequência da checagem de fatos, exclusão das respostas de busca (*shadowban*); d) desmonetização da conta e imposição de multas financeiras; e) outros, como o sistema do Youtube de aplicação de sanções cumulativas, que implica na suspensão da conta.

As tipologias teóricas de Gillespie (2019), Goldman (2021) e Grimmelman (2015) sugerem como os métodos de governança nas plataformas são amplos e variados. O Youtube é um exemplo dessa atuação multifacetada com diversos mecanismos de moderação. A empresa publicou em 2019 o documento “The Four Rs of Responsibility” detalhando os procedimentos de governança do conteúdo considerado danoso. A plataforma estipula quatro dimensões de medidas: *remove* (remover), *raise* (amplificar), *reward* (recompensar) e *reduce* (reduzir). Segundo a plataforma, a criação das regras de moderação é relativa a cada contexto, explicando que, em casos mais problemáticos, o desenvolvimento e revisão das políticas pode levar diversos meses, envolvendo a consulta de especialistas externos ao Youtube, criadores de conteúdo e especificidades regionais (YOUTUBE, 2019a). Análise comparativa dos termos de serviços das plataformas *mainstream* e alternativas indicou que o Youtube possui regras mais inteligíveis, claras, granulares e ilustradas em relação às demais plataformas (SINGHAL et al., 2022).

A exclusão ocorre em três níveis: a) canais: a conta e todos os seus vídeos são removidos de uma só vez; b) vídeos: análise realizada com base no conteúdo ou nos metadados de título e descrição; e c) comentários escritos pelos usuários. A remoção é apresentada pelo Youtube como a principal sanção contra conteúdos que violam as guias de comunidade, na medida em que a velocidade de exclusão e a redução de tempo de exposição de vídeos que não estão em conformidade com as normas são as métricas centrais do relatório de transparência da empresa. Vídeos aos quais são aplicados esse tipo de política frequentemente são disseminados em plataformas sombrias, com menor grau de moderação (ALVES, 2021).

Já a monetização é introduzida como uma medida positiva para recompensar a criação de conteúdo classificado pela empresa como confiável e de autoridade. Dessa forma, fica claro o poder das plataformas que agem como infraestruturas que classificam e organizam informações, atores e comportamentos, traçando linhas simbólicas e concretas entre o que é aceitável e inaceitável (BOWKER; STAR, 2000). O exercício de classificação se transforma em sanção quando o Youtube revoga esses privilégios e desmonetiza um canal em função da violação das regras da comunidade (CAPLAN; GILLESPIE, 2020). Por outro lado, a recompensa pode se tornar um mecanismo de incentivo ao espalhamento da desinformação e discurso de ódio em casos de aplicação inconsistente da política ou de julgamentos incorretos ou enviesados, como foi o caso da Jovem Pan que, a despeito de ser um importante expoente de comunicação das redes de direita, frequentemente produzindo desinformação e discurso de ódio (SANTINI et al., 2023), recebeu milhões de reais no Youtube pelo programa Google News Initiative e foi desmonetizada somente no fim de novembro de 2022.

As duas últimas medidas representam as principais ações de *soft moderation*: amplificação e demerção/redução. A amplificação (*upranking*) é o efeito de destaque no ranqueamento do mecanismo de busca e no algoritmo de recomendação de vídeos. Esse efeito de expansão algorítmica de visibilidade é aplicado para canais que são classificados pelo Youtube como fontes detentoras de credibilidade em temáticas “notícia, ciência e eventos históricos” (YOUTUBE, 2019b, s/p.). Há dois processos epistêmicos importantes de serem notados. O primeiro é a compreensão de que não são todas as temáticas que são passíveis de amplificação, com a implicação de uma seleção entre categorias que são privilegiadas em detrimento de outras; e o segundo é o poder de definição da plataforma de qual tipo de fonte é categorizada como detentora de “credibilidade”. Eventuais problemas nessa classificação também podem transformar uma recompensa em incentivo algorítmico à visibilidade de desinformação, particularmente

considerando que a Jovem Pan e a Fox News foram listadas pelo próprio Youtube como fontes de confiança: “Em 2017, começamos a priorizar vozes autorizadas, incluindo fontes de notícias como CNN, *Fox News*, *Jovem Pan*, India Today e The Guardian, para buscas de notícias e informações nos resultados de pesquisa e nos painéis ‘assistir em seguida’” (YOUTUBE, 2019, s/p. *itálicos adicionados*).

Por fim, a redução de visibilidade (*downranking*) é uma sanção utilizada para conteúdos que são classificados como limítrofes (*borderline*), isto é, são percebidos como problemáticos pela plataforma, mas não chegam a quebrar termos específicos da política de moderação de conteúdo do Youtube. A companhia cita a má-informação danosa (*harmful misinformation*), que é classificada com o auxílio de checadores externos e especialistas (YOUTUBE, 2019b). Segundo o Youtube, as decisões de consenso passam a ser utilizadas como base de dados para treinar algoritmos de aprendizado de máquina que irão automatizar a identificação de vídeos limítrofes e reduzir seu potencial de alcance.

## Procedimentos metodológicos

Os estudos aplicados sobre moderação de conteúdo em plataformas digitais enfrentam dois desafios: a falta de transparência pública e a dificuldade de acesso aos dados. As informações sobre a aplicação das medidas de governança não estão disponíveis nas interfaces programáticas das plataformas digitais. Dessa forma, as pesquisas vêm elaborando métodos para reconstruir digitalmente os metadados da remoção. Particularmente, De Keulenaar e Rogers (no prelo) argumentam em favor de uma mudança na perspectiva teórica dos métodos digitais sobre a análise de traços em plataformas digitais. Segundo o argumento dos autores, pesquisas sobre o digital frequentemente tratam os dados como informações “brutas” e “sem obstrução” da opinião pública. Ao contrário, De Keulenaar e Rogers enfatizam a artificialidade dos dados digitais em diversas dimensões, especificamente, em relação às medidas de governança que jogam luz sobre efeitos sociotécnicos da agência infraestrutural das plataformas. Nesse sentido, trata-se de analisar, prioritariamente, os efeitos da plataforma na modulação dos fenômenos e das mensagens como objeto e locus de investigação.

Para isso, a proposta dos autores é desenvolver métodos de “arquivamento dinâmico”, sistemático e contínuo das plataformas digitais a fim de “reconstruir as cenas” de momentos anteriores e posteriores à aplicação de medidas de moderação de conteúdo. De um lado, isso envolve utilizar mecanismos como o *Wayback Machine* para traçar as transformações nas diretrizes de moderação do Youtube; e, de outro, fazer engenharia reversa dos traços de moderação por meio da combinação de técnicas de extração de dados. Essa abordagem enfatiza oportunidades de estudo num contexto pós-APIs<sup>2</sup> (BRUNS, 2019) e de restrições das políticas de cessão de dados pelas plataformas (D’ANDRÉA, 2021), na medida em que constrói bases de dados que não estão disponíveis de outra forma.

## Lista de canais

Os trabalhos relacionados compõem amostras por meio de termos de consulta (*query design*) a fim de arquivar publicações sobre determinada temática ou *hashtag*, como a covid-19 ou a #stopthesteal (DE KEULENAAR; BURTON; KISJES, 2021; SUZOR, 2020). Todavia, o problema desse desenho de pesquisa é que as táticas de ofuscação da moderação, particularmente o neologismo lexical (DE KEULENAAR e ROGERS, no prelo), dificultam significativamente a tarefa de identificação de vídeos que visam escapar à detecção e, portanto, são mais difíceis de capturar na coleta, como aqueles que usam “vachina”, “v4cina”, “v4c1na” e derivados. Seguindo abordagem de Rauchfleisch e Kaiser (2021), realizamos uma amostra de 849 canais no Youtube que postaram vídeos sobre temáticas político-eleitorais em 2022. A lista é resultado de um processo de amostragem por bola de neve feito pelo módulo *Channel Network* da ferramenta *Youtube Data Tools*, que analisa como canais se seguem e recomendam uns aos outros. Fizemos duas rodadas de expansão da rede em grau de profundidade 1 com o controle da verificação manual a partir de dois critérios: a) idioma em português do Brasil; b) ter publicado ao menos um vídeo sobre o contexto político-eleitoral brasileiro no ano.

## Composição do banco de dados por arquivamento dinâmico

Em seguida, criamos um sistema hospedado em nuvem com um *script* escrito pelo autor utilizando o pacote *tubeR* da linguagem de programação R para acessar a API do Youtube e coletar os últimos 10 a 40 vídeos da lista de canais a cada seis horas, entre 15 de julho de 2022 e 15 de janeiro de 2023. Os metadados foram consolidados eliminando duplicatas, resultando em uma base de dados de 193.429 vídeos. Na primeira semana de fevereiro de

<sup>2</sup> Termo utilizado para contextualizar um momento em que as *Application Programming Interfaces* (APIs) públicas, tradicionais fontes de dados para pesquisadores, estão fechando, como os casos de Instagram, Facebook e Twitter.

2023, realizamos novas consultas à API do Youtube para atualizar as métricas de todas as publicações. Pela interface programática, os vídeos sem estatísticas estão indisponíveis no site (DE KEULENAAR; BURTON; KISJES, 2021; SUZOR, 2020). Criamos um segundo *script* para controlar um navegador em sessão anônima<sup>3</sup> com apagamento de *cookies* a cada sessão para acessar todas as *urls* de vídeos indisponíveis e raspar o texto das razões informadas pelo Youtube.

Como há diferentes motivos para deletar um vídeo, muitos com frequência muito baixa (SUZOR, 2020), criamos as seguintes categorias para agrupar as mensagens de remoção:

- Moderado pelo autor: mensagens avisando que os próprios canais tornaram os vídeos privados ou deletaram o conteúdo;
- Violação: mensagens informando que o vídeo violou uma regra de comunidade ou termo de serviço do Youtube;
- Conta encerrada: mensagens que dizem respeito à deplataformização do canal como um todo e apagamento dos vídeos em consequência;
- Direitos autorais: violação da política de direitos autorais;
- Indefinido<sup>4</sup>: mensagem padrão “*this video is unavailable*” sem maiores detalhamentos.

### **As políticas de moderação de conteúdo**

Analisamos as alterações pertinentes à eleição brasileira entre 2022 e 2023 realizadas no texto da página *Election Misinformation Policies* do Youtube. Complementamos as informações com os dados do relatório “YouTube e as Eleições Brasileiras de 2022: Retrospectiva” elaborado pela plataforma para dar transparência aos esforços de moderação em âmbito da cooperação com o Tribunal Superior Eleitoral (TSE) para combate à desinformação.

### **Resultados**

Neste trecho, apresentaremos os resultados das análises realizadas, organizadas em quatro subtítulos: 1) as políticas de moderação de conteúdo no Brasil; 2) quais foram as razões indicadas pelo Youtube; 3) séries temporais da exclusão; e 4) as proporções da moderação por por canais.

### **As mudanças nas políticas de integridade eleitoral do Youtube**

O primeiro aspecto a se ressaltar é que o Youtube não tinha uma política pronta para o contexto brasileiro em 2022 e foi constituindo ou mudando as regras durante o primeiro semestre ou mesmo durante o pleito eleitoral. As políticas de moderação de conteúdo sobre eleição focam em três categorias: a) políticas de práticas enganosas; b) violência, discurso de ódio e assédio; e c) falsificação da identidade e falsa interação. Especificamente, a primeira regra proíbe conteúdo manipulado, supressão do voto, informação falsa sobre elegibilidade de candidatos, incitação à interferência em processos democráticos (filas, quebra de segurança eletrônica) e declaração de que as eleições foram fraudadas ou que falhas alteraram os resultados do pleito.

Todavia, a aplicação das regras é feita somente depois da certificação institucional do resultado eleitoral. Em 21 de abril de 2022, a política foi atualizada para adicionar a cláusula de integridade eleitoral contra afirmações de fraudes às eleições de 2018. Em 10 de agosto, essa mesma norma foi aplicada também para a eleição de 2014. Enquanto que somente em 31 de outubro, dia posterior à divulgação dos resultados do segundo turno pelo TSE, o Youtube inseriu o pleito de 2022 expressamente na política de integridade eleitoral.

A plataforma possui uma regra de acúmulo de sanções em que os canais que receberem três avisos de remoção de conteúdo por violação dos termos de serviço dentro de 90 dias serão banidos permanentemente. Contudo, as alterações nas políticas de integridade eleitoral foram seguidas por períodos de tolerância de 30 dias em que os vídeos que quebrassem os termos poderiam ser excluídos do site, mas não contariam para a suspensão permanente. De fato, como foram três mudanças em datas diferentes, não foram aplicadas sanções cumulativas durante noventa dias, inclusive no mês posterior à eleição. “Isso significa que conteúdos que violaram nossas diretrizes postados entre

3 Seguimos boas práticas da literature especializada para testes autônomos e sistemáticos de navegadores. A automação foi realizada pelo Selenium WebDriver (<https://www.selenium.dev/>) e operacionalizada num ambiente virtual Docker (<https://www.docker.com/>) para fins de estabilidade e replicabilidade.

4 Consideramos a categoria “Indefinido” como uma exclusão feita pela plataforma sem especificar a razão. Entendemos que não é uma ação tomada pelo próprio canal, o que está contemplado em outras mensagens. Todavia, não encontramos documentação pública informando claramente como interpretar estes casos e seguimos estudos relacionados. Este é um desafio importante, também criticado pela literatura especializada.

31 de outubro de 2022 e 30 de novembro de 2022 foram retirados de nossa plataforma, mas os canais não tiveram outra penalização além da eliminação desses vídeos” (YOUTUBE, 2023, p.3).

Cabe questionar por que a corporação tomou decisões distintas para as disputas eleitorais de 2014 e 2018, o que duplicou o tempo de tolerância para 60 dias num ano central para a democracia brasileira. Ademais, chama atenção que o reconhecimento da lisura da eleição de 2022 tenha sido incluído na política somente depois do segundo turno, o que permitiu um período de não aplicabilidade da política durante todo o ano, já que, até a promulgação do resultado, os termos só faziam referência à 2014 e 2018.

O Youtube indicou no relatório de transparência que mais de 10 mil vídeos e 2,5 mil canais foram removidos por terem violado as políticas sobre as eleições no Brasil. Segundo os dados da empresa, mais de 84% foram excluídos antes de alcançar 100 visualizações. Contudo, para além de esforços de arquivamento dinâmico como o realizado nesta pesquisa, não há como obter esse banco de dados e verificar ou avaliar de forma independente o relatório.

### Razões expressas pelo Youtube

Encontramos 29.755 vídeos indisponíveis, representando 15,4% de uma amostra de 193.429; totalizando 870,4 milhões de visualizações em média de 29,2 mil por unidade de análise. Considerando as *urls* que estavam offline, apresentamos as razões informadas pelo Youtube na página (TAB.1).

**Tabela 1** - Status dos vídeos, razão e visualizações

Status	Frequência	Visualizações (total)	Visualizações (média)
Online	163.690 (84.6%)	5.644.636.495	34.483.7
Indefinido	17.384 (9.0%)	485.843.722	27.947.75
Canal	11.769 (6.1%)	371.879.172	31.598.2
Conta encerrada	421 (0.2%)	5.796.027	13.767.29
Violação	142 (0.1%)	4.395.405	30.953.56
Copyright	22 (0.0%)	2.474.471	112.476

**Fonte:** Elaborada pelo autor com base em dados extraídos do Youtube.

Os resultados mostram que a 9% das remoções apresentavam apenas a mensagem “*This video isn't available anymore*” sem maiores detalhamentos ou justificativas; outros 6,1% dizem respeito às ações dos próprios canais, considerando que 7.709 vídeos foram tornados privados e 4.060 foram deletados. Esses valores são importantes e apontam para a exclusão não como consequência da ação das plataformas, mas como um ato preventivo das próprias contas.

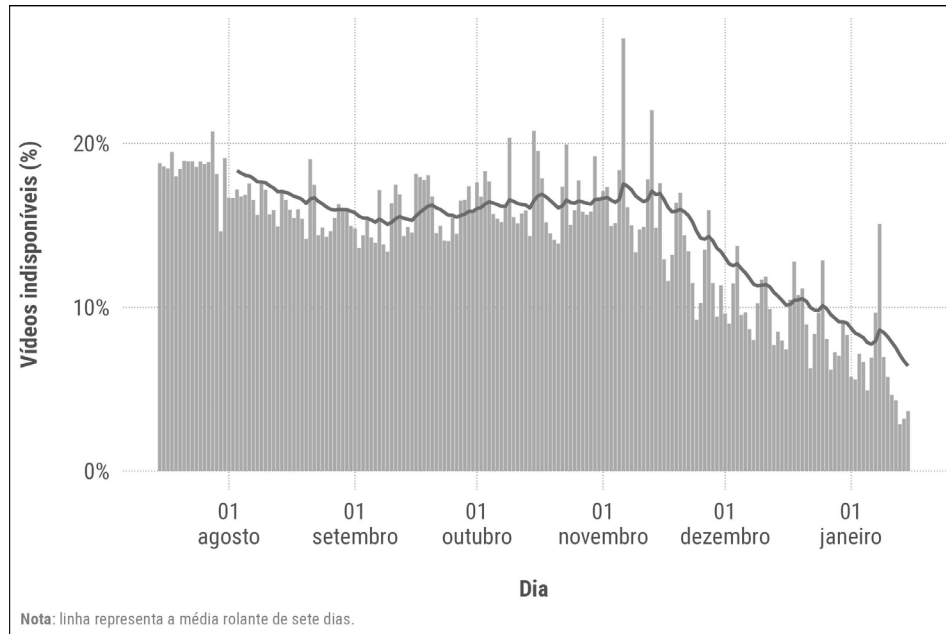
Vale ressaltar que apenas 142 vídeos desativados contém justificativa de violação de alguma política do Youtube. Todavia, a falta de transparência permanece, na medida em que 118 casos tem uma mensagem genérica, como: “*This video has been removed for violating YouTube's Community Guidelines*”, em outros 15 está escrito “*This video has been removed for violating YouTube's Terms of Service*”. As duas justificativas mais frequentes que indicam expressamente qual regra foi quebrada são: “*This video has been removed for violating YouTube's policy on harassment and bullying*”, cinco vezes, e “*This video has been removed for violating YouTube's policy on violent or graphic content*”, quatro vezes. Destaca-se que nenhum vídeo tem mensagem explicando a exclusão em função de quebra da política de integridade eleitoral ou de discurso de ódio, o que impede análise detalhada e revisão dos critérios de moderação da plataforma.

Levando em conta a subamostra que possui mensagem de violação explícita, o vídeo com maior alcance foi “ACABOU de ACONTECER 30.10 #rio” do canal #MANOTOKIO2020, removido por quebrar a diretriz de conteúdo violento ou explícito, que, em nosso último registro, tinha atingido 1,1 milhões de visualizações. Chama a atenção que mesmo os vídeos com temática eleitoral identificada no título não tenham recebido informação compatível com a exclusão, como: “FA x TSE - Teremos eleições [sic] confiáveis?” postado pelo Brasil pela Direita, “BRAZIL WAS STOLEN - Auditoria Privada das Eleições 2022” de Tramonte ou “Auditores argentinos atestam fraudes nas eleições brasileiras” de Carlos Ferrari.

### Dinâmica temporal da remoção

Outro aspecto a ser analisado é quando foram postados os vídeos que estão indisponíveis. Em tese, isso poderia quando se concentram os episódios de quebra nas diretrizes ou alterações no investimento de recursos para moderação por parte da plataforma. A Figura 1 demonstra qual o percentual diário de publicações indisponíveis no momento da verificação em fevereiro de 2023.

**Gráfico 1:** Percentual de vídeos indisponíveis por dia e linha da média rolante de sete dias.

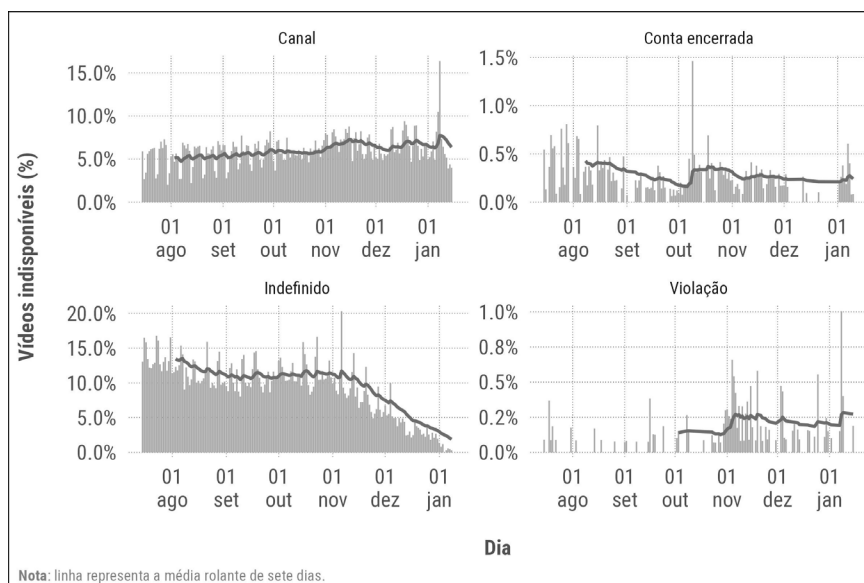


**Fonte:** Elaborada pelo autor com base em dados extraídos do Youtube.

Os resultados da série temporal apontam que julho foi o mês com maior proporção de vídeos removidos (18,66%). Há uma estabilidade entre agosto (16%), setembro (15,5%) e outubro (16,6%). Apesar de ser o mês da votação de primeiro e segundo turno, momento em que, hipoteticamente, as mensagens seriam mais extremas, não houve grande crescimento de remoções em outubro. Aconteceram dois picos: 06 de novembro com 29,1% da amostra *offline* e 13 de novembro com 23,1%. Ao todo, em dez datas, mais de 20% dos vídeos estão indisponíveis.

O principal achado da análise de série temporal é a queda contínua e expressiva que ocorre depois da votação de segundo turno. Somente 6,24% dos conteúdos estavam *offline* em janeiro. Podemos levantar três hipóteses concorrentes para interpretar este dado: 1) com o resultado das eleições, os canais reduziram postagens de conteúdo político-eleitoral ou aprenderam a ofuscar as mensagens que questionavam mais frontalmente a vitória de Lula e a lisura do processo eleitoral; 2) embora a plataforma tente remover violações rapidamente, esse procedimento é muito demorado; ou 3) o Youtube reduziu a dedicação de esforços recursos, equipe e sistemas de moderação de conteúdo que foram acionados para atuar durante os meses da campanha. A Figura 2 desagrega as razões dos vídeos indisponíveis para jogar luz sobre essas questões.

**Figura 2:** Total de vídeos indisponíveis por dia considerando a razão da remoção.



**Fonte:** Elaborada pelo autor com base em dados extraídos do Youtube.



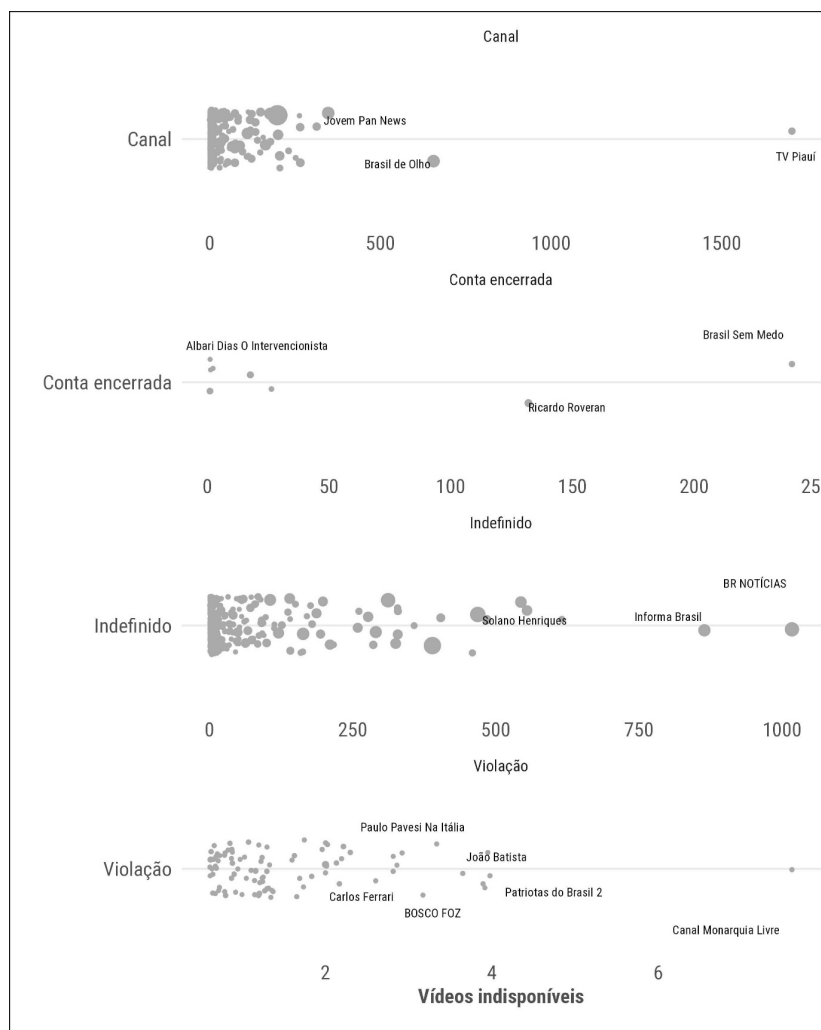
Quando se observa separadamente os motivos das remoções, percebe-se uma inversão das tendências. Durante todo o período eleitoral, o volume de conteúdos removidos por ação do Youtube é mais elevado. Esse padrão começa a mudar a partir do dia seguinte à votação de segundo turno. De fato, do dia 1 de novembro em diante, há uma queda constante do rótulo “Indefinido” e uma manutenção com pequeno decréscimo das medidas tomadas pela gestão dos canais. Da segunda semana de dezembro em diante, as linhas se invertem e atingem maior discrepância no dia 08 de janeiro, marco da depredação dos prédios dos Três Poderes em Brasília, em que 163 vídeos foram apagados por quem postou, ao passo que apenas 10 foram removidos pelo Youtube por violação das regras, quatro sem razões específicas e três em função do encerramento da conta. Isso indica que os canais continuaram postando mensagens que, potencialmente, desafiavam as regras do Youtube e tomaram medidas para gerenciar o risco a posteriori, deletando ou tornando o conteúdo privado, enquanto a plataforma relaxou as medidas de moderação.

**A exclusão de vídeos por canais**

Alguns canais foram ser completamente apagados, o que inviabiliza o acesso a todos os vídeos produzidos. Considerando a amostra desta pesquisa, 49 contas (5,7%) tiveram 100% do conteúdo indisponível no momento da verificação. Isso quer dizer que se um estudo começasse a coletar dados de fevereiro de 2023 em diante, pelo menos 49 canais foram apagados da base de dados do Youtube e estão inacessíveis para investigação a posteriori. De forma mais geral, 116 canais (15%) têm mais de 75% dos vídeos offline.

Quando se observa o detalhe dos dados, nota-se que alguns canais apagaram quase todas as publicações realizadas durante o período dessa pesquisa. A TV Piauí, por exemplo, deletou ou tornou privado 1705 vídeos (99,88%) da amostra coletada. O mesmo padrão se repete para Patriota em Ação (99,12%), Pastora Valdirene Moreira (99,1%), Renato R Gomes (98,33%), TV Bolsonaro Rumo à 2022 (98,05%) e Brasil de Olho (97,47%).

**Figura 3:** Totais de vídeos removidos no Youtube agrupados por canal



Fonte: Elaborada pelo autor com base em dados extraídos do Youtube.

Já as remoções por violação são menos comuns. A conta com mais exclusões nessa categoria foi o Canal Monarquia Livre com oito violações (0,51%). No total, 13 canais atingiram a marca de três violações no período estudado por este artigo. Todavia, não se pode avaliar mais claramente a política de aplicação de *strikes* que implicam em bloqueio permanente do canal em função dos períodos de tolerância e do prazo de 90 dias.

## Conclusões

Esse artigo analisou a aplicação de medidas de moderação de conteúdo no Youtube entre 15 de julho de 2022 e 15 de janeiro de 2023. A partir de debates teórico-metodológicos que repropõem o estudo de traços digitais para elucidar os efeitos e agências das plataformas (DE KEULENAAR e ROGERS, no prelo), o desenho de pesquisa reconstitui as cenas da remoção de vídeos por meio do arquivamento dinâmico de metadados e raspagem das razões da indisponibilidade de conteúdos. A pesquisa pretende contribuir com o marco teórico sobre governança de plataformas digitais, particularmente acerca da moderação de conteúdo e desinformação político-eleitoral (GILLESPIE, 2017; GOLDMAN, 2021).

A bibliografia especializada sobre a governança nas plataformas discute criticamente o exercício de poder privado dessas corporações tecnológicas como editoras ativas dos discursos na esfera pública, sem procedimentos constituídos democraticamente, legitimação pública ou adequação às especificidades nacionais (SCHARLACH; HALLINAN; SHIFMAN, 2023). Embora as políticas de moderação do Youtube sejam uma das mais inteligíveis (SINGHAL, et al, 2021), indicamos que a maioria dos vídeos é removida sem informar publicamente as razões e as políticas que foram quebradas, o que reduz a transparência e *accountability* do arcabouço de governança. As medidas de implementadas pelas plataformas (GILLESPIE, 2019) devem ser complementadas por legislação debatida com a sociedade civil e aprovada pelo Congresso Nacional, a fim de regulamentar as práticas de remoção de conteúdo (KLONICK, 2017). Além disso, em linha com estudos sobre a dimensão temporal da governança (SUZOR, 2020) enfatizamos a relevância do período pós-eleitoral em eleições contenciosas, demonstrando que o Youtube reduziu os esforços de moderação de conteúdo durante o período de intensa contestação dos resultados da eleição e mobilizações em frente aos quartéis militares.

Os resultados ressaltam que é extremamente desafiador pesquisar contextos de acirramento político-eleitoral com grandes riscos à integridade democrática na plataforma. Há um problema de resgate da memória digital e acessibilidade a dados para investigações sobre movimentos antidemocráticos, desinformação e violência política, na medida em que quase um terço dos vídeos deixam de existir no Youtube, apenas um mês depois do fenômeno (RAUCHFLEISCH; KAISER, 2021). Além disso, os achados reforçam a necessidade de aprovar regulamentações e mecanismos procedimentais para ampliar a transparência pública sobre as decisões de moderação tomadas pelas plataformas e preservação posterior da base de dados para a academia (HARTMANN et al., 2023; SUZOR, 2020).

A governança de plataformas é uma agenda de investigação relevante para os estudos de mídia, tecnologia, direito e ciências sociais. Estudos futuros podem analisar mais sistematicamente o conteúdo dos vídeos removidos em comparação com os que ainda estão online. Parte da análise deve questionar a consistência da aplicação das políticas de moderação da plataforma. Todavia, é fundamental problematizar esses regramentos tendo em vista conceitos normativos da teoria democrática ou do arcabouço conceitual dos direitos humanos, a fim de avançar na proposição de modelos de governança e normas de comunidade mais participativos e representativos de valores substantivos e contextos nacionais.

## Referências

- ALVES, Marcelo. Clones do YouTube: replataformização da irrealdade 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. 146144448231156580, 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. 146144448211045662, 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.