Moderation of Political Content on Youtube during the 2024 US Election

Andreu Casas*

Abstract

Today social media platforms play a crucial role in moderating political information and speech. Yet, despite growing concerns, we still know very little about how often nor the conditions under which platforms moderate political content. For one year leading to the 2024 US election, we monitored the moderation of content from more than 11,000 salient YouTube channels posting about US politics. The platform removed 1.8% of the 3.2 million videos that we tracked. Suspension rates were higher among videos about politics (2.6%, *v.* 1.6% for non-political content such as sports), among political videos on Foreign Trade (10%), Defense (4.4%) and Government Operations (3.3%), and videos posted by conservative channels (3.2% *v.* 1.3% for liberal channels). **[to be completed]**

*Royal Holloway University of London. Department of Politics, International Relations and Philosophy: andreu.casas@rhul.ac.uk. This research has received funding from a VENI grant from NWO (VI.Veni.211R.052, PI: Andreu Casas).

Introduction

An increasing number of people rely on social media for political engaement. For example, about half of the US population get news from social media platforms, particularly Facebook (30%) and YouTube (25%) (1). In turn, a handful of private social media companies today have an unprecedented power to regulate political information and speech. A growing number of scholars discuss the implications of such paradigm shift in content moderation for politics and democracy (2; 3), including concerns about commercial (4) and geopolitical incentives (5; 6), governmental pressures (7), and lack of democratic input and due process (2; 8). In addition, claims about potential ideological biases in content moderation proliferate among political elites and the public. As some examples, Republicans in the US argue that major platforms censor conservatives at higher rates (9; 10), and some humanitarian and Palestinian groups complain about social media algorithms favoring pro-Israel content (11). However, despite the academic and societal urgency, due to a lack of transparency and independent research, we still know little about the conditions under which major social media platforms moderate political content. To a large degree, research on this topic is yet scarce due to the technical and methodological complexities associated with collecting, processing, and analyzing large quantities of social media data. The few exceptions that exist mostly focus on Twitter (12; 13; 6) or on small samples of e.g. YouTube videos (10).

Here, we contribute to our understanding of the moderation of political content on social media by monitoring on one of the largest platforms (YouTube), a big number of elite and salient channels posting about (US) politics (∼10,000 channels, for a total of 2.5 million videos), and for a long and relevant period of time (1 year leading to the 2024 US election). Although platforms have many moderation tools in their tool belt, such as search bans or down ranking (12), we focus here on the most drastic form of moderation: channel and video removals. We ask five questions that our crucial to our understanding of political so-

cial media moderation. (RQ1) First, how often does YouTube remove channels and videos of political relevance? Although the platform reports some aggregate numbers (https://transparencyreport.google.com/youtube-policy/removals: e.g. 788,354 video removals in the US from October to December 2023), we do not know about the potential political nature of the sanctioned channels/content. (RQ2) Second, does YouTube remove political (*v.* non-political) content at different rates? Channels that post about politics can also post about other non-political topics (e.g. news channel posting a sports clip). To understand the moderation of political content we need to look at the removal of videos of political nature *vis-a-vis* non-political videos. (RQ3) Third, why are channels and videos of political nature removed? Platforms take down content and accounts for a variety of reasons, such as hateful conduct and misinformation. The large and relevant sample of channels and videos here provide a unique opportunity to explore the distribution of motivations behind the suspension of political content on a major platform. (RQ4) Fourth, are removals equally frequent across political topics – or are videos on a particular topic the target of moderation during this relevant electoral period? (RQ5) Finally, addressing the aforementioned claims about ideological biases in content moderation, are conservative channels (and their videos) more likely to be removed (*v.* liberal channels and their videos)?

Data and Method

Removals happen but they are likely to be rare. In order to be able to explore meaningful variations in content moderation, we wanted to identify and monitor a large and diverse pool of salient YouTube channels posting about US politics. We used a snowballing technique (6; 14). The starting point was an extensive list of YouTube channels of media organizations (e.g. New York Times, Fox News, etc. N = 105) and of US politicians (members of the 118th Congress, President Biden, and former president Trump. N = 184). Then, in April 2023 we identified

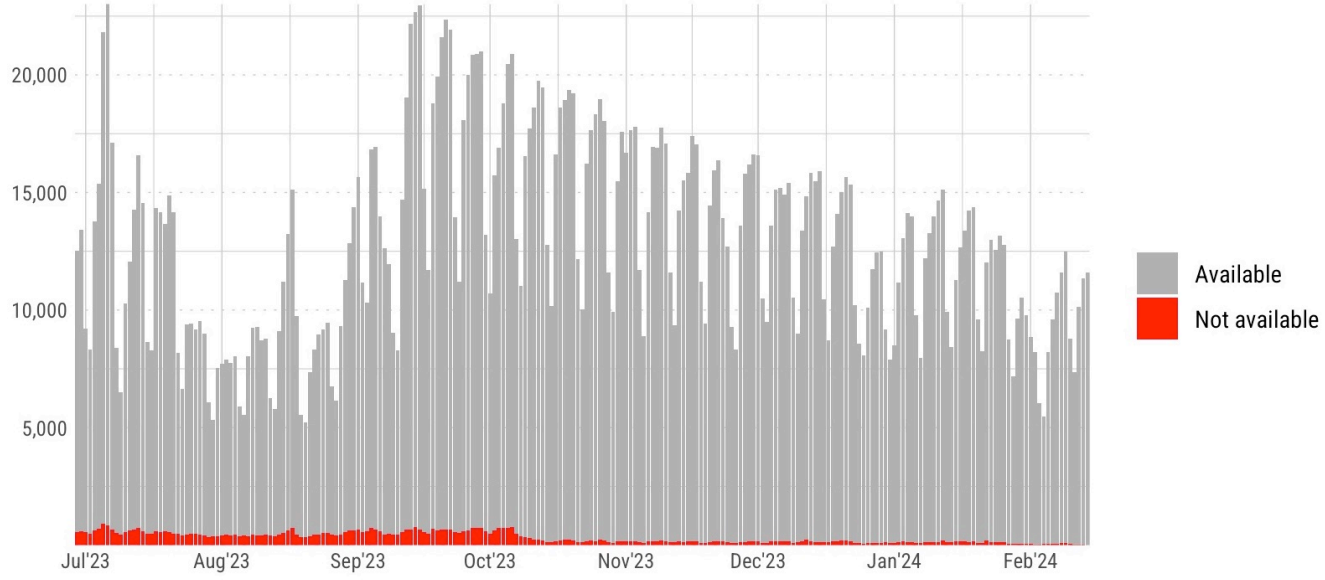
a set of politically-engaged users who commented or replied to videos from elite channels (N = 26,353 users), and collected the full list of channels to which they were subscribed (N = 7,334,005 channels – 1,624,136 of them unique). We narrowed the list to focus on the most relevant channels: those to which at least 10 of our politically-engaged users were subscribed (N = 92,653). Finally, we trained a language model that used channel descriptions and tags to predict whether they posted about politics. To account for the wide range of data sources from which people can get political information, we trained the model to identify channels as political even if only some of the its content was of political relevance. Through this process we identified 20,054 politically-relevant channels (such as Russell Brand, Brian Tyler Cohen, and The Ring of Fire, but also additional media and politician channels) and added them to the original list of elite channels, for a total of 20,343.

On June 28th 2023 we started monitoring these channels using the YouTube API. We developed a set of computer scripts to collect, on a rolling basis (about once a week per channel), the following information: a) whether the channels were still active, b) if not, the reason why they had been removed, c) if active, all videos (metadata, transcript, and unique video frames) posted since the last time we had checked (or a random sample of 100 videos if they had posted >100 since the last check), d) whether the previously collected videos were still active, and e) if not, the reason why they had been removed. In addition, we trained machine learning models to estimate the ideology of the channels (liberal, moderate, or conservative), and to predict which videos were about US politics (*v.* a non-political topic), and if so, the particular political topic (based on the 21 topics of the Comparative Agendas Project, (15)). For ideology we used correspondance analysis, a widely-used and validated dimension-reduction method that can leverage information about audience (subscribers) overlap among elite accounts (media and politician channels) to estimate the ideology of social media users (16; 17; 6). To identify political videos and their topic, we trained language models to perform each task. In *Material and Methods* we provide further details on the data and methodology.

Results

Figure 1 shows the amount of videos collected from June 28th 2023, to February 15th 2024: a total of 3,204,440 videos from 11,016 unique channels, for an average of 12,841 videos a day (95% CI: 12,294-13,389). The volume of videos was substantially lower during the summer months (mid-July to mid-September), and in general there is a downwards trend that is likely to be explained by the gradual removal of channels in the sample. About 3.8% of the videos (N = 121,087) stopped being publicly accessible by the end of data collection – red bar areas in Figure 1. However, at a first glance we do not know if these were removed from the public domain by the platform or the users themselves. The subsequent analyses focus on 10,816 channels for which we were able to estimate an ideology, and on 2,441,233 videos sent by these channels and for which we were able to obtain a transcript and to generate a political and topic prediction.

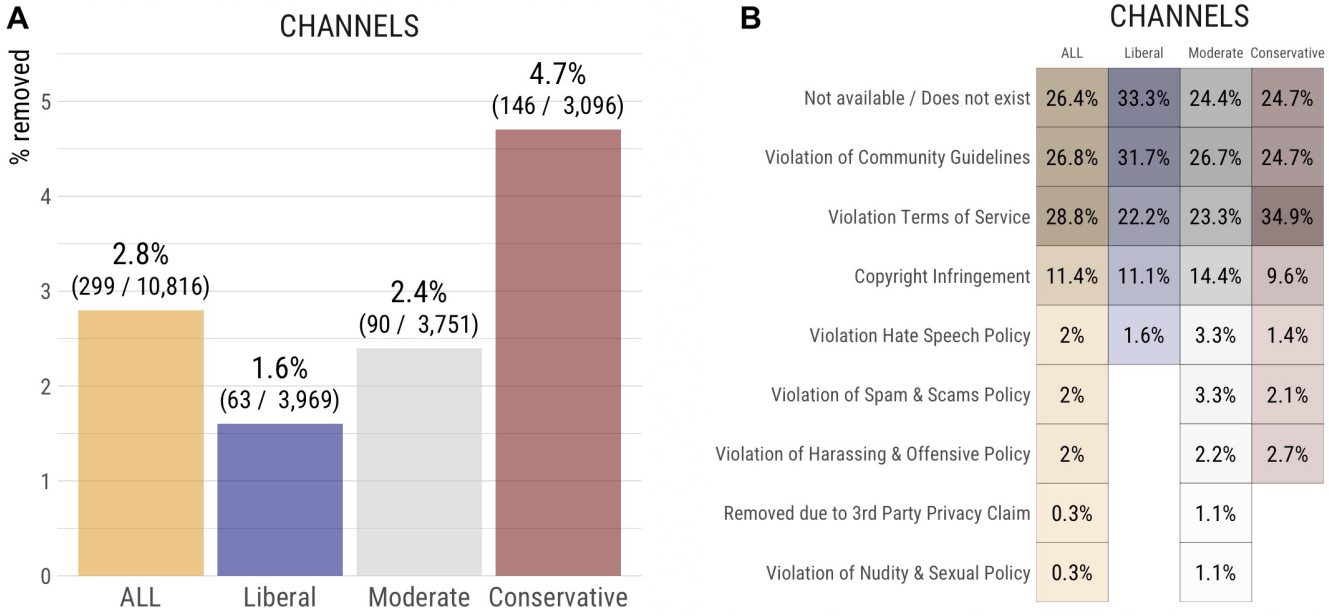
Figure 1: Number of YouTube videos collected from June 28th 2023, and whether they are still available on the platform.



Channel removal

We first look at the removal of channels (RQ1). About 2.8% were removed by Feburary 15th 2024 (299 out of 10,816). Note that we only consider here channels removed by YouTube (3 additional channels had been taken down by the users). As shown in Figure 2.A, there are clear ideological differences (R5). Although we started monitoring a slightly larger amount of liberal channels (N = 3,969; *v.* 3,751 moderate and 3,096 conservative), the suspension rate was substantially lower for this group: 1.6%, compared to 2.4% for moderate channels, and up to 4.7% for conservative ones.

Figure 2: **A.** Percentage of channels removed by YouTube by February 15th 2024 (by ideology of the channel). **B.** Reasons reported by YouTube for removing the channels (% of channels from a given ideology).



In Figure 2.B we look at the reasons given by YouTube for removing these 299 channels (RQ3). The platform does not provide this information via the API. Instead, we scraped the reason provided by YouTube when trying to access the channel via a web browser. We observe again some ideological differences, with liberal channels being suspended at higher rates

for violating the community guidelines (31.7% of liberal channels removed) and conservative channels being removed more frequently for violating the terms of service (34.9%). But more importantly, it is striking the vast ambiguity regarding the removal of most channels. About 82% of all removed channels are simply reported as "Not available", not existing, or as having been removed for general violations of the "Community Guideliens" or the "Terms of Service". Concrete reasons such as "Copyright Infringement" or violations of the "Hate Speech Policy" are provided in relation to the suspension of only about 12% of the channels.

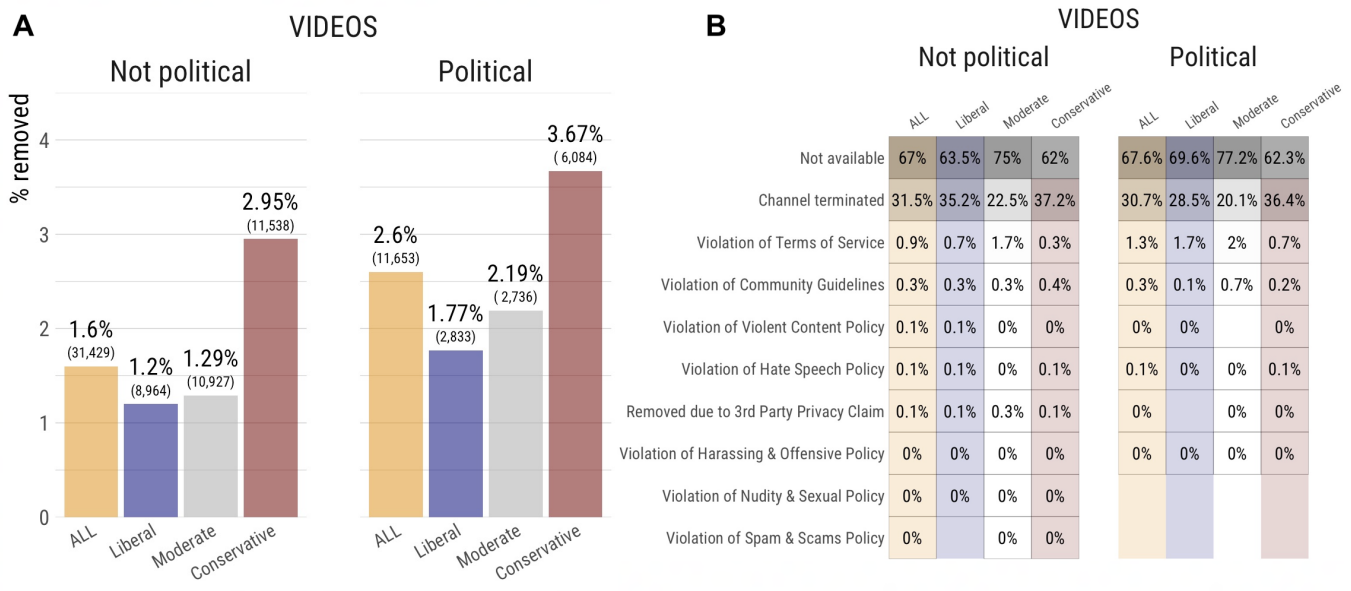
Video removal

A total of 43,082 of the 2,441,233 videos analyzed here (1.8%) were removed by YouTube during the period of analysis (RQ1). These do not include another set of 37,824 videos that also became unavailable because the channels made them private (N = 22,855), removed them (N = 9,093), were taken down only temporarily by YouTube (N = 5,876), or as a result of the host channel being terminated by the user (N = 269).

Figure 3.A shows removal rates by ideology (RQ5) and distinguishing between the political nature of the video (RQ3). Our language model predicted 451,099 of the videos to be about US politics (18.5%). Removal rates are substantially higher (+60%) among these political videos: 2.6% *v.* 1.6% for videos that were not about US politics. In terms of ideology, in Figure 3.A we observe a pattern similar to the one in Figure 2, videos posted by conservative channels are removed at higher rates (3.67% for videos about US politics, and 2.95% for other videos) compared to moderate (2,19%, 1.29%) and particularly to liberal channels (1.77%, 1.2%).

Figure 3.B shows the reasons provided by YouTube regarding the removal of these videos (RQ2). Again, very similar (although even more pronounced) patterns to those shown in Figure 2 regarding the removal of channels emerge. The platform provides very vague statement regarding the suspension of most videos. For 99% of them, the platform simply states that

Figure 3: **A.** Percentage of videos removed by YouTube by Feburary 15th 2024 (by ideology of the channel). **B.** Reasons reported by YouTube for removing the videos (% of videos from a channel from a given ideology).

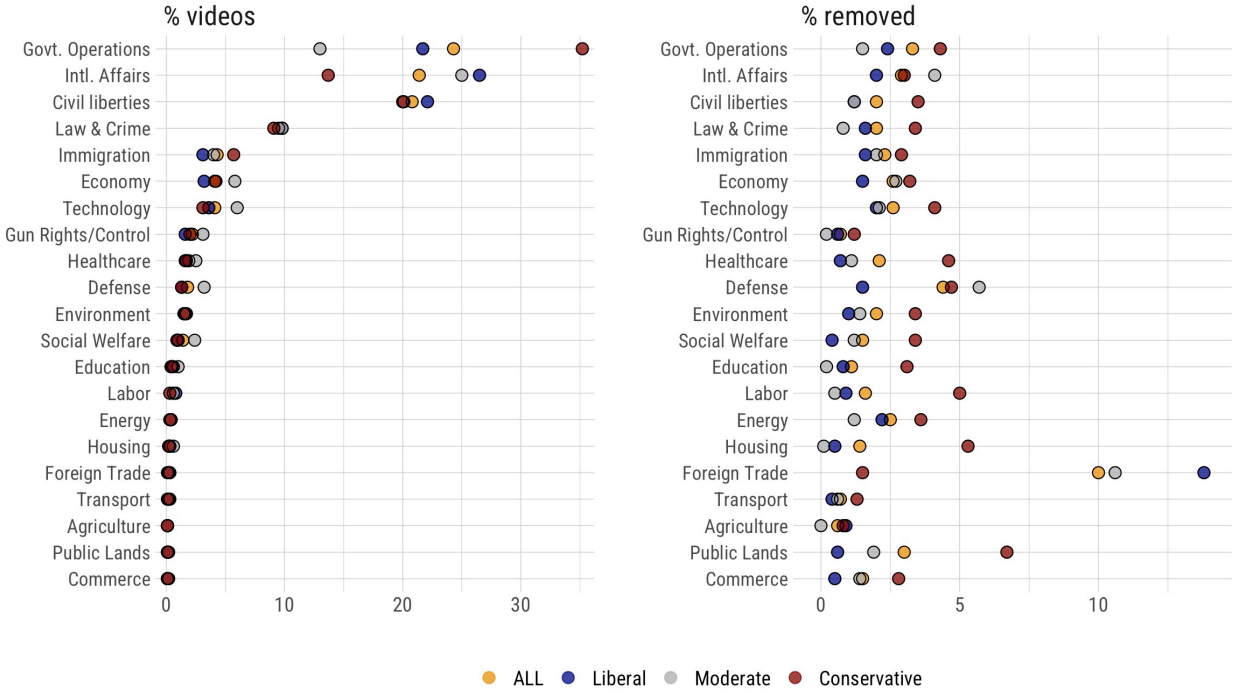


they are "Not available", that the channel has been terminated, or a general violation of the terms of service or community guidelines; and only provide detailed reasons for the remaining 1%, such as violations of the "Violent Content" or "Nudity and Sexual" Policy.

Next we focus only on the political videos (N = 451,099) and look at removal rates across topics (RQ4). Figure 4.A first shows how frequently channels from each ideology posted videos on each of the topics. The most discussed topics were government operations (includes issues related to the bureaucracy, inter-branch relations, but also electoral campaigns and scandals), international affairs, civil liberties and law and crime. Government operations was the most discussed topic among conservative channels, whereas international affairs was the top issue among liberal ones. Topics such as public lands or agriculture were rarely discussed by any channel. When turning our attention to Figure 4.B, we observe similar removal rates across topics: between 0 and 5% approximately. The only exception is Foreign Trade, with

a substantially higher removal rate for liberal (14%) and moderate (11%) channels. However, this was a very marginal topic with very few videos, even for liberal and moderate channels in general. In regards to the other topics, the highest removal rates are for videos on defense (4.4%) and government operations (3.3%). In addition, across the board we observe higher removal rates for videos from conservative channels.

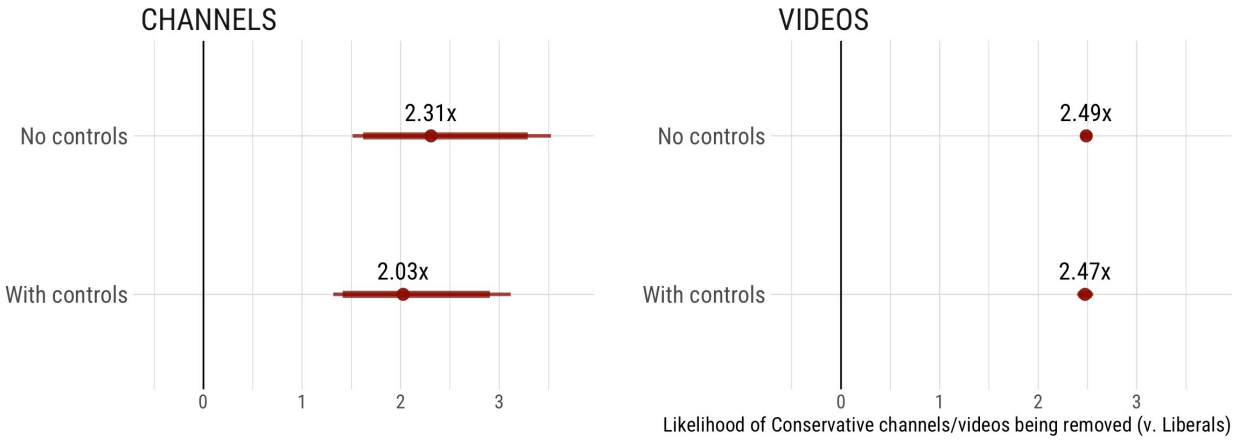
Figure 4: **A.** Percentage of political videos on each topic, by ideology. **B.** Percentage of removed political videos on each topic, by ideology.



Ideological differences

The previous analyses consistently showed higher removal rates for conservative channels and their content, compared to liberal and moderate channels and videos. Many Republican elites complain about a pro-liberal moderation bias in major social media platforms (9), but such ideological differences could also be explained by conservative channels posting at higher rates

Figure 5: Likelihood of Conservative channels/videos being removed (v. Liberal channels/videos). These difference in likelihoods come from four logistic regressions predicting removal.



content that violates some of YouTube's community guidelines – such as using hate speech or spreading misinformation. The lack of clarity on YouTube's side regarding the removal of channels and videos, as shown in Figures 2.B and 3.B, makes it difficult to assess. Figure 5 shows a preliminary attempt at doing so. It shows results from four logistic regressions, two predicting channel suspension (left panel) and two others predicting the removal of videos (right panel). The first set of regressions predict the removal of channels/videos only as a function of the ideology of the channel (*No controls*). The other two regressions predict the same outcomes but we also included a range of covariates (*With controls*). Some of these covariates come from a pretrained language model (`detoxify`) that we used to predict the likelihood of the videos in our sample to be: toxic, obscene, attack someone's identity, insult, threat, and contain explicit sexual content. We averaged these video-level predictions at the channel level for the model predicting channel removal. Moreover, in this channel-level model we also include the number of overall videos, and the number of political videos sent during the period of analysis (both logged), as additional controls. And in the model predicting video

removals we also included the topic of the video as an additional control.

Figure 5 shows that even when accounting for these additional explanations, conservative channels and videos are removed more often than liberal ones. In both cases, the ideological difference is smaller in the models with controls, but conservative channels still remain 2.03 times more likely to be removed than liberal ones (from 2.31 times in the model without controls); and conservative videos still are 2.47 times more likely to be removed (down from 2.49 in the model with no controls).

Discussion

[**to be completed**]

Material and Methods

[**to be completed**]

Supplementary Materials

Supplemental material for this article is available [**to be completed**].

References

[1] E. Shearer, A. Mitchell, Social media and news fact sheet, *Pew Research Center* (2023).

[2] J. M. Balkin, Free speech in the algorithmic society: big data, private governance, and new school speech regulation, *UCDL Rev.* **51**, 1149–1210 (2017).

[3] T. Gillespie, *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media* (Yale University Press, 2018).

[4] S. T. Roberts, *Behind the screen: Content Moderation in the Shadows of Social Media* (Yale University Press, 2019).

[5] J. Earl, T. V. Maher, J. Pan, The digital repression of social movements, protest, and activism: A synthetic review, *Science Advances* **8**, eabl8198 (2022).

[6] A. Casas, The geopolitics of deplatforming: A study of suspensions of politically-interested iranian accounts on twitter, *Political Communication* **0**, 1-22 (2024).

[7] R. Gorwa, *The Politics of Platform Regulation: How Governments Shape Online Content Moderation* (Cambridge University Press, 2024).

[8] J. C. York, *Silicon values: The future of free speech under surveillance capitalism* (Verso Books, 2022).

[9] J. Davalos, B. Brody, Facebook, twitter ceos sought by senate over n.y. post story., *Bloomberg:* https://www.bloomberg.com/news/articles/2020-10-15/facebook-twitter-chided-anew-by-republicans-over-ny-post-story (2020).

[10] S. Jiang, R. E. Robertson, C. Wilson, *Proceedings of the International AAAI Conference on Web and social media* (2019), vol. 13, pp. 278–289.

[11] H. R. Watch, Metas broken promises systemic censorship of palestine content on instagram and facebook, *Human Rights Watch* (2023).

[12] K. Jaidka, S. Mukerjee, Y. Lelkes, Silenced on social media: the gatekeeping functions of shadowbans in the American Twitterverse, *Journal of Communication* **73**, 163-178 (2023).

[13] S. Majo-Vazquez, M. Congosto, T. Nicholls, R. K. Nielsen, The role of suspended accounts in political discussion on social media: Analysis of the 2017 french, uk and german elections, *Social Media + Society* **7**, 20563051211027202 (2021).

[14] P. Barberá, *et al.*, Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data, *American Political Science Review* **113**, 883–901 (2019).

[15] B. D. Jones, *et al.*, Policy agendas project: Codebook (2023).

[16] P. Barbera, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, Tweeting from left to right: Is online political communication more than an echo chamber?, *Psychological Science* **26**, 1531-1542 (2015). PMID: 26297377.

[17] M. Wojcieszak, A. Casas, X. Yu, J. Nagler, J. A. Tucker, Most users do not follow political elites on twitter; those who do show overwhelming preferences for ideological congruity, *Science Advances* **8**, eabn9418 (2022).

Appendix A Estimating the ideology of YouTube channels.

Figure A1: Validation of the CA-based ideology scores