Paying Attention to the Algorithm Behind the Curtain

Bringing Transparency to YouTube's Demonetization Algorithms

ARUN DUNNA, Yext Inc. & University of Massachusetts - Amherst
KATHERINE A. KEITH, Allen Institute for Artificial Intelligence
ETHAN ZUCKERMAN, University of Massachusetts - Amherst
NARSEO VALLINA-RODRIGUEZ, IMDEA
BRENDAN O'CONNOR, University of Massachusetts - Amherst
RISHAB NITHYANAND, The University of Iowa

YouTube has long been a top-choice destination for independent video content creators to share their work. A large part of YouTube's appeal is owed to its practice of sharing advertising revenue with qualifying content creators through the YouTube Partner Program (YPP). In recent years, changes to the monetization policies and the introduction of algorithmic systems for making monetization decisions have been a source of controversy and tension between content creators and the platform. There have been numerous accusations suggesting that the underlying monetization algorithms engage in preferential treatment of larger channels and effectively censor minority voices by demonetizing their content.

In this paper, we conduct a measurement of the YouTube monetization algorithms. We begin by measuring the incidence rates of different monetization decisions and the time taken to reach them. Next, we analyze the relationships between video content, channel popularity and these decisions. Finally, we explore the relationship between demonetization and a channel's view growth rate. Taken all together, our work suggests that demonetization after a video is publicly listed is not a common occurrence, the characteristics of the process are associated with channel size and (in unexplainable ways) video topic, and demonetization appears to have a harsh influence on the growth rate of smaller channels. We also highlight the challenges associated with conducting large-scale algorithm audits such as ours and make an argument for more transparency in algorithmic decision-making.

1 INTRODUCTION

YouTube, with over 2B active monthly users and 52M content creators [86, 96], dominates the online video-sharing marketplace. In addition to its role in democratizing access to audiences and creative tools, the popularity of YouTube among independent media creators is owed in large part to its early and longstanding practice of sharing advertising revenue with *"advertiser friendly"* content contributors through the YouTube Partner Program [97]. As the revenue collected by content creators has continued to grow,[1] becoming a "YouTuber" or an independent media creator is now seen as a popular and viable career option [40, 50, 62, 63]. However, content creators' increasing dependence on payments from YouTube as a source of revenue makes the (algorithmic) determinations of what constitutes "advertiser friendly content" – i.e., monetizable content – critical.

While YouTube's reliance on machine learning algorithms for (de)monetization decisions is widely accepted by creators due to the scale of the classification task at hand, the opacity of the algorithm has been a source of controversy. The opacity of YouTube's demonetization algorithm has led to many claims of unfair treatment by content creators. Most notably the algorithm was at the center of numerous lawsuits filed by LGBTQ [3], Black [85], and conservative [67] YouTubers alleging biases in YouTube's recommendation and demonetization algorithms against content

[1]The period between 2019 and 2020 saw a 50% and 40% increase in the number of creators earning over $10K/year and $100K/year, respectively. [96]

Authors' addresses: Arun Dunna, dunna@pm.me, Yext Inc. & University of Massachusetts - Amherst; Katherine A. Keith, katherinek@allenai.org, Allen Institute for Artificial Intelligence; Ethan Zuckerman, ezuckerman@umass.edu, University of Massachusetts - Amherst; Narseo Vallina-Rodriguez, narseo@imdea.org, IMDEA; Brendan O'Connor, brenocon@cs.umass.edu, University of Massachusetts - Amherst; Rishab Nithyanand, rishab-nithyanand@uiowa.edu, The University of Iowa.

curated specifically for their audiences. Further allegations from content creators have suggested YouTube employs differential and favorable monetization treatment for “premium” partners with large numbers of subscribers [39, 89]. Despite the growing list of accusations and mounting anecdotal evidence of bias, there has not been a large-scale systematic study of the demonetization algorithm used by YouTube. In this work, we seek to fill this gap in knowledge by studying the characteristics of YouTube’s algorithm-driven demonetization process. Specifically, we seek to find answers to the following research questions:

- **RQ1.** *What are the incidence rates of (de)monetization decisions made by YouTube?* (§4). We use a heuristic and longitudinal data to understand the frequency of and characteristics associated with demonetization (i.e., monetized → non-monetized) and remonetization (i.e., monetized → non-monetized → monetized) decisions. Our analysis shows that only a small fraction of videos (0.48%) appear to experience these specific transitions in monetization status. However, for these videos, the median time to a final monetization status is nearly 5 days and the median time to complete a remonetization transition is 13 hours longer than for demonetization transitions – suggesting the cost of the human review process associated with remonetization decisions.
- **RQ2.** *How is the monetization process associated with channel and video characteristics?* (§5) We categorize videos by popularity and content. We then condition the demonetization and remonetization rates on each of these features to bring transparency to some of the triggers of demonetization and remonetization. Our analysis finds that in comparison to more popular channels, less popular channels experience higher demonetization rates through a faster process and lower remonetization rates through a slower process. We also discover that certain topics are in fact subject to higher rates of demonetization and remonetization. By manual inspection, we see that many of these decisions are explainable (e.g., copyright violations) while several are not.
- **RQ3.** *How is demonetization associated with a channel’s future growth rate?* (§6). Drawing from a difference-in-differences analysis design, we estimate the expected growth rate of channel views to understand the effects that demonetization decisions have on content creators. Our analysis shows an estimated average effect of −11.8 percentage point loss in channel views for creators of demonetized videos, with the influence being most strongly felt by channels with between 100K and 1M subscribers (−30.3 percentage points).

Taken together, our investigation yields insights into the workings of YouTube’s opaque demonetization algorithm with a focus on its accuracy, triggers, and effects on content creators. Our results provide support to concerns regarding preferential treatment for larger channels and unexplainability of the demonetization process. In §8 we highlight the challenges and limitations faced by auditing studies such as ours and call for more transparency in algorithmic decision-making.

2 BACKGROUND: YOUTUBE MONETIZATION

In this section, we provide a high-level overview of YouTube’s publicly available demonetization policies and processes in place during the period in which our data was gathered. Since our analysis is focused on data gathered between July and September 2020, we only highlight the policies in place prior to September 2020. These policies relevant to the period of our study are also available through the Internet Archive’s WayBack Machine [6–13].

Becoming a YouTube Partner. To qualify for monetization, content creators are required to join the YouTube Partner Program (YPP). This requires their channels to meet the following criteria:

- *Viewership criteria.* Creators need to have, at a minimum, 4K watch hours in the past 12 months and 1K channel subscribers.

- *Content criteria.* Creators need to agree to follow YouTube's community guidelines [12] which are aimed at curbing: promotion of tools for engagement metric inflation and ad fraud, impersonation of individuals and channels, spam and scams, harassment and cyberbullying, dangerous and violent content, hate speech, nudity and sexual content, and the sale and manufacturing of firearms and illegal goods.
- *Review criteria.* A final automated and human review is then conducted to verify that the channel does not violate community content guidelines. This review focuses on the channel theme, latest and most popular videos, and metadata (e.g., thumbnails and video descriptions).

Creators satisfying the above criteria may then choose to monetize individual videos.

Video monetization. Not all YPP-member created videos may be monetized. Monetization is only possible on content meeting YouTube's advertiser-friendly content guidelines [6]. These guidelines impose further restrictions on the use of profanity, violent, graphic, sexually suggestive, demeaning, tobacco-related, and controversial or sensitive content in videos or their metadata. Importantly, the initial process of identifying whether a video meets the advertiser-friendly guidelines is completely automated via the demonetization algorithm. YouTube provides two options for creators seeking to monetize their content during the video upload process [8]:

- *Upload as unlisted.* Videos may be uploaded as unlisted, making them accessible to the demonetization algorithm and unreachable to the general public. An initial algorithmic monetization decision is made available to the creator within 20 to 60 minutes of this upload. Creators may then decide to either make an appeal for human review, publicly list, or edit their content and seek another automated decision.
- *Self-certify content.* Videos made by creators with a history of producing advertiser-friendly content may be immediately monetized if their creators self-certify that their video meets YouTube's monetization guidelines. However, videos may be demonetized and the creators ability to self-certify content may be restricted if later checks find violations of the advertiser-friendly content policy. Repeated violations may also result in exclusion from the YouTube Partners Program.

Video demonetization and appeal. It is important to note that the initial decision made by the demonetization algorithm at upload time need not be final. Automated decisions are algorithmically re-evaluated and may change multiple times within the first 48 hours of the upload. Additionally, the nature of viewer engagement with the uploaded content may cause further changes to the video monetization status even after the first 48 hours [7]. Therefore, videos may go through cycles of demonetization and remonetization during their lifetime – with particularly high frequency within the first 48 hours of their upload. Notification of demonetization decisions are communicated via a 'yellow dollar' icon placed by the relevant video on the creators' dashboard. These automated demonetization decisions may then be appealed – effectively asking for human review of the algorithm's decision. Human review of a video may take between a few hours and weeks. Of note is YouTube's statement that reviews of videos getting substantial traffic are a priority for human reviewers [7]. This suggests that the oft-complained about preferential treatment towards larger channels is by design.

Channel strikes, demonetization, termination, and appeals. Channels producing content found to be in violation of the community guidelines are subject to the following enforcement policy [10]: First, the violating content is removed from YouTube and a notice of violation is issued to the creator. Creators making their first violations are issued an official warning with no further action. Violations made after this warning result in a 'strike' against a channel. Channels receiving more than three strikes in a 90-day period are terminated. As with the video demonetization process, the initial decisions to issue strikes against channels are made by automated algorithms and appeals of these decisions can be made which result in human review. Further, although specifics are unavailable, prior to channel termination YouTube may also resort to monetization-related actions including demonetizing an entire channel and suspending creator participation in the YouTube Partner Program in the event of community guideline violations [18].

Updates since September 2020. We note that there have been numerous updates to YouTube's content guidelines, advertising, and monetization policies since the time of our data gathering [15, 97]. Most notably, in November 2020 a policy change allowed ads to be shown even on content not monetized by the creator [20, 44]. We discuss the significance of these updates and their impact on future studies such as ours in §8.

3 THE YOUTUBE DATASET

In this section, we provide an overview of our data collection process (§3.1-§3.2), ethical considerations (§3.3), and characteristics of our dataset (§3.4). At a high-level, our data collection pipeline involved two simultaneously running processes – a video ingestion process (§3.1) and a monetization status monitoring process (§3.2). The video ingestion process sought to identify popular and/or controversial videos uploaded by YPP-member channels. The monetization status monitoring process regularly recorded the monetization status of all videos selected by the video ingestion process.

3.1 Video ingestion

The primary goal of our video ingestion process is to develop a dataset of videos that: (1) contain sufficient controversial content so that the behaviors of the demonetization process can be observed and (2) is reasonably representative of YPP-member created content on YouTube and *not just content recommended by YouTube*. We explicitly make this distinction because of the possibility that content recommended by YouTube (e.g., via the 'Trending' tab or on the front-page) is already deemed advertiser-friendly as a result of interaction between the YouTube recommendation algorithm and monetization algorithm – as has been suggested in prior work [61]. Consequently, we populated our dataset using a combination of external sources (Reddit) and YouTube recommendations. We do this using the modules described below. The interactions between each of these modules are also illustrated in Figure 1.

Controversial channel identifier. In order to develop a sample of controversial channels, we begin by scraping YouTube links from all the comments and posts made to 326 subreddits obtained via the Pushshift dataset [29]. These 326 subreddits include 128 popular subreddits associated with politics (obtained the related subreddits page on /r/politics) and 198 subreddits which were banned or quarantined by Reddit prior to July 2020. This latter group are subreddits that Reddit has determined have violated their community guidelines (e.g., antivax content, hateful content, etc.); we use this as a proxy to identify controversial communities, topics, and potential links to YouTube content that may also violate YouTube's own monetization and community guidelines. We note that this dataset specifically aims to oversample controversial content to allow for a systematic study of YouTube's demonetization process. Reddit was chosen as an external source for this controversial content for several reasons including data availability, public announcements of administrative subreddit bans and quarantines, and popularity. This module outputs the set of channels associated with all the YouTube links found in our 326 subreddits and adds them to the *monitored channels* list. In the remainder of this paper, where appropriate, we analyze videos made by channels identified through this mechanism as a separate group. Doing so allows us to study the demonetization process experienced by controversial content creators.

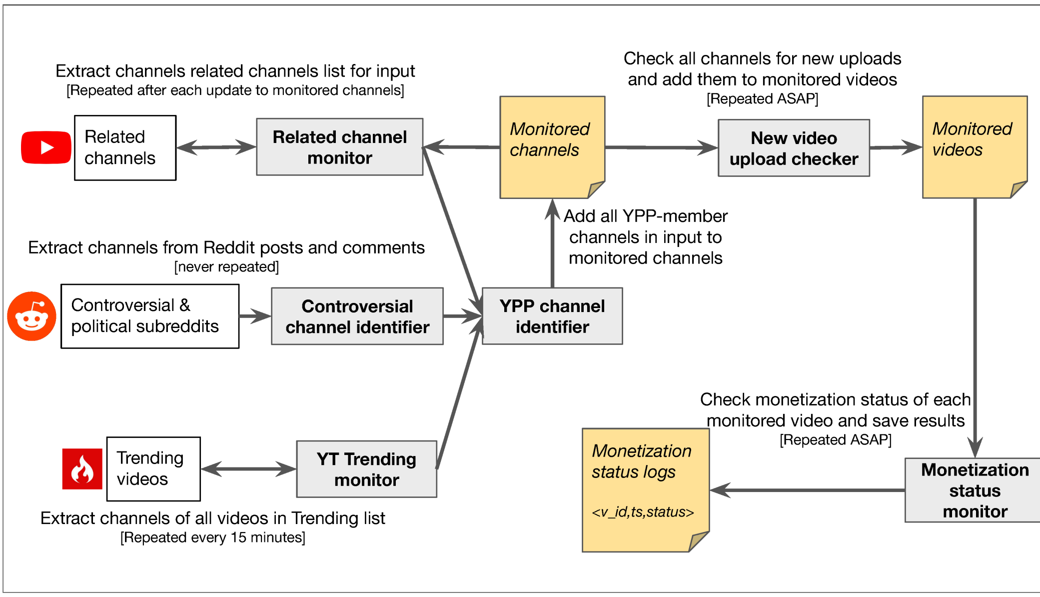


Fig. 1. Architecture of the data collection pipeline for generating our video monetization status dataset. Transparent boxes represent video ingestion sources, grey boxes represent the modules of our data collection framework, and yellow boxes indicate the lists maintained by our framework.

YouTube Trending monitor. Via the 'Trending' tab, YouTube provides a regularly updated feed of videos that are likely to be appealing to a large number of viewers. Videos displayed in this tab are not personalized (i.e., they are common to all viewers) and are selected based on the rate at which views are gathered by the video, amongst other signals. Importantly, YouTube maintains specific safety filters, which includes human staff, to vet the content and appropriateness of videos selected to appear on the trending tab. Updates to the Trending videos are made every fifteen minutes by YouTube [95]. Following each update, this module outputs the set of channels associated with the current YouTube trending videos. This set is added to the *monitored channels* list. In the remainder of this paper, where appropriate, we analyze the videos made by channels identified through this mechanism as a separate group. Doing so allows us to obtain a view of the demonetization process experienced by creators of mainstream content.

Related channel monitor. Channel creators may also provide a set of related channels on their YouTube channel page. These are accessible through the 'channels' tab on the channel homepage. This module takes a list of channels as input and outputs a set of channels marked as related to the input channels. This module is called after each addition update to the *monitored channels* list and its output is added to this list.

YPP channel identifier. This module takes a list of channels as input and prunes all channels that are not a part of the YPP program. This is done by removing all channels where: (1) the channel had less than 1K subscribers or (2) no ads were displayed on any of their prior uploads. Note that at the time of this study, YouTube did not show ads on non-monetized content — a policy that changed in November 2020 [20]. The pruned list is returned as the output.

New video upload checker. Using the most current *monitored channels* list as input, this module uses the YouTube API [19] to check for the existence of newly created content from any of the channels in the '*monitored channels list*'. All newly uploaded videos identified in this step are added to our *monitored videos* list. We note that these checks occurred at least once every two hours during the period of our study from July 22 to September 9, 2020. Therefore, the upper-bound between the time a video was publicly accessible and the time that it was added to our *monitored videos* list is two hours.

3.2 Monetization status monitoring

Our method is based on the fact that, prior to November 2020, YouTube showed ads only on YPP-member monetized content. Therefore, the observation of variables associated with requesting and rendering ads in the pages of a loaded video was sufficient to infer monetization status. Specifically, our **monetization status monitor** module takes a video from the *monitored video list* as input simply checks for the presence of the `adTimeOffset` and `playerAds` variables in the HTML of the video. The `adTimeOffset` dictionary indicates the times at which an ad will be played, and the absence of this dictionary in the loaded page indicates that no ads will be played. Similarly, the `playerAds` variable contains a list of dictionaries related to rendering options for loaded ads and is absent when no ads are to be loaded. In order to ensure valid results, we conduct repeated measurements to correctly infer monetization status. Specifically, each video is loaded five times in succession and each of the obtained pages is checked for the presence of the above variables. A video is classified as 'non-monetized' only if these variables are absent in all five pages. The module outputs a tuple indicating the video ID, timestamp, and monetization status.

We validated our method, during the time of data gathering, by confirming that these variables were not present on non-monetized content. This was done by confirming the absence of these tags on content from creators not qualifying for the YPP.

Obtaining temporal monetization status data. Since we are interested in monetization status transitions, it is important to obtain multiple monetization status measurements for each video in our dataset. Further, our analysis on the time taken for these transitions necessitated a high frequency of measurements for each video. To satisfy these needs, we run the monetization monitoring module with a minimum-delay round-robin approach described below.

- *Minimum delays between consecutive videos.* The absence of a delay between successive videos results in IP-blocking by YouTube due to suspicions of malicious traffic. To avoid this, we limit our rate of video monetization status checks to a maximum of 1 video/second/IP. Our measurement infrastructure allowed us to use 50 IPs resulting a rate of 50 status checks/second. We note that a request for an IP-blocking exemption was declined.
- *Zero delays between consecutive measurement iterations.* One measurement iteration is a monetization status check on all videos in the *monitored videos list*. No delays were added between two measurement iterations. This approach results in the smallest achievable interval between consecutive monetization status checks for each video. A consequence of this design decision is that the time interval between status checks for a video is variable and increases as our dataset grows. For our measurement infrastructure, the interval between the start of two measurement iterations was always under two hours – i.e., each video had a status check conducted at least once every two hours. This implies that when a monetization transition is observed, the actual time of its occurrence is never greater than two hours prior to the time recorded by our system. Thus our time-to-transition results serve as upper-bounds.

Limitations of our inferences. Our method for inferring monetization status suffers from limitations. First, our inability to observe transitions in monetization state as soon as they happen may result in incorrect categorization of the video. For example, a video may transition from monetized → non-monetized → monetized between two consecutive measurement iterations and our system incorrectly labels the video as 'always monetized'. However, this particular limitation only causes over-counting of 'always monetized' and 'never monetized' videos (the 'No Transition' group in Table 2) while leaving no false-positives in our set of videos labeled as having experienced a monetization status transition (the 'Transition' group in Table 2). Second, our method is unable to distinguish between videos that are kept non-monetized by their creators and those that are algorithmically non-monetized. We argue that our results are still representative since: (1) creators are unlikely to leave content non-monetized when it is monetizable; (2) creator decisions to voluntarily leave content non-monetized might be indicative of the creators own conclusion that the content is not advertiser-friendly and therefore likely to trigger algorithmic demonetization (we discuss this possibility in the context of related work in §7); and (3) we break down our analysis in a way that allows us to focus on cases where explicit transitions in monetization status, which are likely to be algorithmically induced, are observed. We note that both limitations are a result of the lack of access to concrete monetization-related signals provided by YouTube and our inability to find alternate signals.

3.3 Ethics of data gathering and release

Data gathering. Our research was facilitated using tools to automatically gather and analyze the metadata associated with YouTube videos and save information relevant to our study. In an effort to mitigate any harmful effects of our study on YouTube, we made the following restrictions:

- *Time period.* We only gathered data for a limited time period from July 22 to September 9, 2020. All data gathered by our tools were explicitly for the purpose of this study.
- *Using the API.* We leveraged the official YouTube API whenever the data sought was accessible through the API. This was for all cases except measurement of a video's monetization status. We explicitly obeyed all rate-limits imposed by YouTube during use of the official API.
- *Scraping limitations.* Our scraping tool was used only for measuring the monetization status of a video which was not available through the API. This scraping did not violate the `robots.txt` restrictions set by YouTube. Further, we implicitly obeyed any rate-limits by throttling our measurements until our traffic was no longer classified as 'suspicious' by YouTube.
- *Only public data was accessed.* Our tools only obtained data that was already publicly available through webpage source code or the YouTube API.

Our use of a scraper is in line with prior auditing studies which have used scrapers to uncover algorithmic discrimination and bias [34, 35, 48] when API access was not sufficient to conduct the audit. Such methods of auditing are categorized as a *scraping audit* by Sandvig et al. [83] in their classification of research methods for algorithm audits. Although their work (written in 2014) highlights the challenges of navigating the Computer Fraud and Abuse Act (CFAA) when leveraging this method, we note that a recent (June 2021) ruling by the Supreme Court in *Van Buren vs. United States* specifically ruled that such scraping, even if found to violate the terms of service of a website, is not a violation of the CFAA. Further details may be found in the Supreme Court opinion [17], a press release from the Electronic Frontier Foundation [16], and the amicus brief filed by several Internet measurement researchers and the American Civil Liberties Union [14].

Data release. In order to facilitate further analysis and research, we intend to release our datasets of video metadata and their associated monetization status at `anonymized`. This release does not include any Personally Identifiable Information (PII), instead only containing timestamped monetization statuses and timestamped number of views for all video IDs in our dataset. We note that the random identifiers assigned by YouTube as video IDs can be used in conjunction with the YouTube API to gather other metadata associated with our videos and their corresponding channels. We use this approach because it allows YouTube creators to exercise access control over

	Total	Trending	Reddit	Related channels
Channels	9,965	5,253	2,646	1,796
Videos	354,884	248,944	69,626	36,314

Table 1. Channels observed by each source and the number of videos created by them and monitored by us during the period of our study. If a channel is seen in multiple sources (e.g., on Reddit and YouTube Trending), we attribute its addition to our dataset only to the source where it was first observed.

the metadata of videos that were turned private or deleted after our initial data collection (since the YouTube API does not respond to queries for metadata of private or deleted videos and channels). Since we do not provide a means to track individual users across sites or provide data that is inaccessible to the public, our approach satisfies suggested guidelines for the use of social media data laid out by Rivers and Lewis [79].

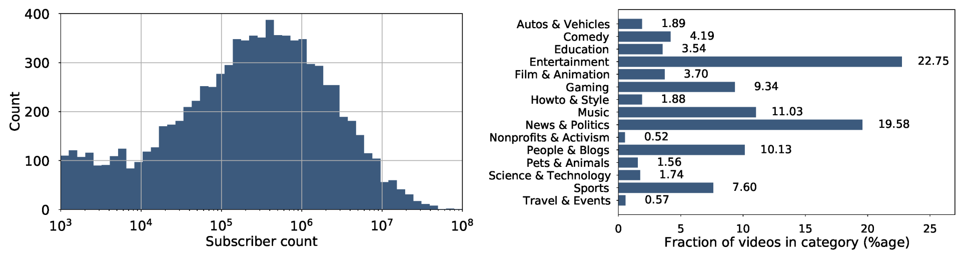
3.4 Dataset characteristics

Our dataset consists of 354,884 videos published by 9,695 channels between July 22 and September 9, 2020. Table 1 shows the number of videos gathered from each of our ingestion sources. We find a majority of our videos were obtained from channels first observed in 'YouTube Trending' videos. This is expected since YouTube: (1) makes updates to these lists every 15 minutes, thus giving our dataset exposure to a large number of new channels and (2) focuses on showcasing content from creators who primarily create content for YouTube [95] which suggests a higher upload rate from the channels observed on Trending. Figure 2a shows the distribution of channel subscriber counts for all channels in our dataset. The mean, median, and 10^{th} quantile subscriber count for channels in our dataset were 1.41M, 269K, and 6.4K, respectively. This high variance in channel subscriber counts allows us to analyze the relationship between channel popularity and demonetization decisions. Figure 2b shows the fraction of videos in our dataset belonging to each category (note that categories are self-selected by creators at the time of upload). There are no publicly available sources of distributions of content from YPP-members against which we may compare our dataset for representativeness. However, in comparison to the general distribution of YouTube content (i.e., including non YPP-member created content) made available by Statista from 2018 [2], our dataset is over-represented in the Entertainment and News & Politics categories while being under-represented in the People & Blogs and Gaming categories. We note that it remains unclear: (1) if our dataset is not representative of YPP-member created content; and (2) what impact any mismatches in content distribution may have on the generality of our findings.

4 INCIDENCE RATES OF DEMONETIZATION

In this section, we focus on answering the following question: **RQ1. What are the incidence rates of (de)monetization decisions made by YouTube?** More specifically, we identify the number of demonetization (*monetized* → *non-monetized*) and remonetization (*monetized* → *non-monetized* → *monetized*) transitions observed for videos in our dataset.

Incidence rates of monetization status transitions. A summary of the measured incidence rates is provided in Table 2. Using the approach outlined in §3.2, we identified 47,949 videos (13.3%) that experienced non-monetization at some point during our study. Of these, only 7,098 videos (1.96% of our dataset) experienced transitions in monetization status – i.e., monetization (non-monetized → monetized), demonetization (monetized → non-monetized), or remonetization (monetized → non-monetized → monetized) transitions. In the remainder of our analysis, we



(a) Channel subscriber counts

(b) Video categories

Fig. 2. Distributions of channel subscriber counts and video categories for the 9,695 channels and 354,884 videos in the dataset, respectively.

Group	Label	Description	#Videos (%)
No Transition	$\mathfrak{D}_{11}$	Always monetized	311,906 (86.67%)
	$\mathfrak{D}_{00}$	Never monetized	40,851 (11.35%)
Transition	$\mathfrak{D}_{10}$	Demonetization: Monetized → Non-monetized	1,242 (0.35%)
	$\mathfrak{D}_{101}$	Remonetization: Monetized → Non-monetized → Monetized	484 (0.13%)
	$\mathfrak{D}_{01}$	Non-monetized → Monetized	5,288 (1.46%)
	$\mathfrak{D}_{multiple}$	Multiple demonetization and remonetization transitions	84 (0.02%)

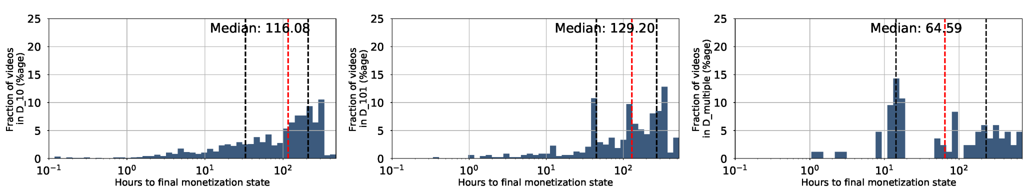
Table 2. Breakdown of videos' observed monetization status, including ones experiencing status transitions, in our dataset.

exclude the $\mathfrak{D}_{01}$ dataset due to our inability to determine if the initial non-monetization decision was made by the creator or by the algorithm. An incorrect assumption here has the potential to lead to misunderstandings of the monetization algorithms. Instead, we focus explicitly on the cases where a transition from monetized to non-monetized was observed after the video was publicly listed – i.e., cases of demonetization ($\mathfrak{D}_{10}$), remonetization ($\mathfrak{D}_{101}$), and multiple transitions ($\mathfrak{D}_{multiple}$. Due to the low likelihood of demonetization transitions being caused by creator decisions, we use these videos to better understand YouTube's monetization process. We explore these further in §5.

- *Demonetized videos* ($\mathfrak{D}_{10}$). The 1,242 videos identified to have experienced only a demonetization transition provide insights into the YouTube algorithmic process for demonetizing content. We use this dataset to provide insights into the speed and content triggers for these algorithmic-decisions.
- *Remonetized videos* ($\mathfrak{D}_{101}$). The 484 videos identified to have experienced a remonetization transition provide insights into the content triggers that cause incorrect decisions by YouTube's demonetization algorithms. Further, since YouTube uses human reviewers to make final decisions on remonetization, these videos also allow us to understand how channel popularity relates to the frequency and speed of remonetization decisions.
- *Multiple transitions* ($\mathfrak{D}_{multiple}$). The 84 videos identified to have experienced multiple demonetization and remonetization transitions. Although they have the potential to provide us with insights into the content triggers that cause disagreements and confusion between the algorithms driving the demonetization process, we find that their utility is limited by the small number of observed instances.

Group	Ingestion source	$\mathfrak{D}_{11}$	$\mathfrak{D}_{00}$	$\mathfrak{D}_{10}$	$\mathfrak{D}_{101}$	$\mathfrak{D}_{\text{multiple}}$
$\mathbf{G}_{\text{trending}}$	Trending	230,699 (92.73%)	16,700 (6.71%)	907 (0.35%)	414 (0.13%)	60 (0.02%)
	Related (Trending)	596 (90.99%)	53 (8.09%)	3 (0.45%)	2 (0.30%)	1 (0.15%)
	Total in $\mathbf{G}_{\text{trending}}$	231,295 (92.72%)	16,753 (6.71%)	910 (0.36%)	416 (0.16%)	61 (0.02%)
$\mathbf{G}_{\text{reddit}}$	Reddit	52,218 (75.14%)	17,047 (24.53%)	153 (0.22%)	47 (0.06%)	21 (0.03%)
	Related (Reddit)	28,393 (79.65%)	7,051 (19.78%)	179 (0.50%)	21 (0.05%)	2 (0.01%)
	Total in $\mathbf{G}_{\text{reddit}}$	80,611 (76.67%)	24,098 (22.92%)	332 (0.31%)	68 (0.06%)	23 (0.02%)

Table 3. Monetization transitions broken down by video ingestion source. Related (Trending) and Related (Reddit) refer to the videos ingested from channels found in the "Related channels" pages of channels observed on YouTube Trending and Reddit, respectively. We group videos ingested from channels observed on YouTube Trending and their related channels into $\mathbf{G}_{\text{trending}}$. Similarly, videos ingested from channels observed on Reddit and their related channels into $\mathbf{G}_{\text{reddit}}$.



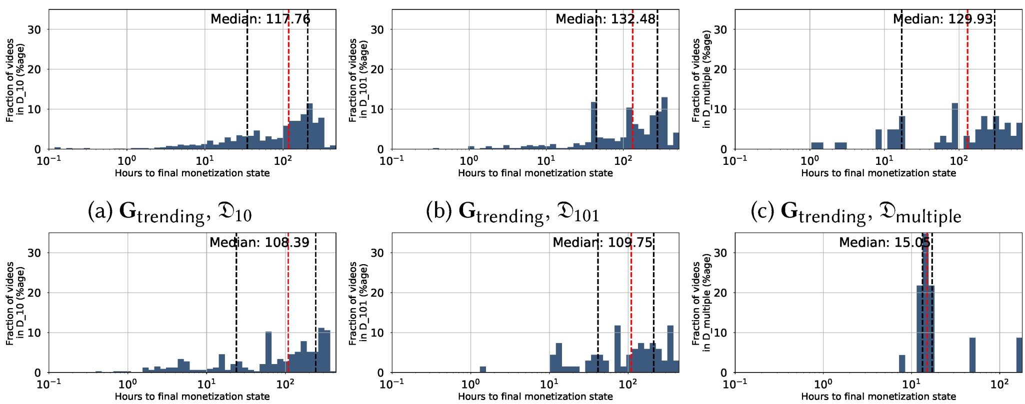
(a) $\mathfrak{D}_{10}$: Demonetized (b) $\mathfrak{D}_{101}$: Remonetized (c) $\mathfrak{D}_{\text{multiple}}$: Multiple transitions

Fig. 3. Time taken by videos to reach their final monetization state. Red lines indicate the median and black lines indicate the 25^{th} and 75^{th} percentile.

For videos where our system always observed the same monetization state, we denote their groups $\mathfrak{D}_{11}$ (always monetized) and $\mathfrak{D}_{00}$ (never monetized); these groups are also shown in Table 2.

Table 3 breaks down the types of monetization transitions observed by each video ingestion source. In this table and in the remainder of this paper, we group together channels observed on YouTube Trending together with the channels observed on their "Related channels" pages (group $\mathbf{G}_{\text{trending}}$) and channels observed on Reddit with the channels observed on their "Related channels" pages (group $\mathbf{G}_{\text{reddit}}$). While keeping channels from related creators grouped together, this grouping also allows us to conduct statistical tests which would not be possible otherwise due to the small counts of monetization transitions observed in the Trending and Reddit "Related channels" datasets. Our analysis shows that, in comparison to channels obtained from Reddit ($\mathbf{G}_{\text{reddit}}$), videos from creators that have appeared on YouTube Trending ($\mathbf{G}_{\text{trending}}$) are less likely to be non-monetized (by 11-17%). Statistical analysis via χ^2 tests of independence confirmed (p<.05) that (1) the final observed monetization status (i.e., monetized or non-monetized), (2) the types of monetization status transitions (i.e., $\mathfrak{D}_{10}$, $\mathfrak{D}_{101}$, and $\mathfrak{D}_{\text{multiple}}$) observed, and (3) occurrence of any transitions in monetization status (i.e., belonging to $\mathfrak{D}_{10}$, $\mathfrak{D}_{101}$, or $\mathfrak{D}_{\text{multiple}}$) were all *not independent* of the ingestion group ($\mathbf{G}_{\text{trending}}$ or $\mathbf{G}_{\text{reddit}}$). This suggests a fundamental difference in the monetization experiences of videos in our ingestion groups.

Time taken for transition to final monetization status. Among videos with a demonetization transition, Figure 3 shows the time taken for videos to transition into their final observed monetization state. Considering only the videos in our $\mathfrak{D}_{10}$, $\mathfrak{D}_{101}$, and$\mathfrak{D}_{\text{multiple}}$ groups, the median observed time to final state was 120.1 hours – well above YouTube's documented claim of most



(a) $\mathbf{G}_{\text{trending}}$, $\mathfrak{D}_{10}$ (b) $\mathbf{G}_{\text{trending}}$, $\mathfrak{D}_{101}$ (c) $\mathbf{G}_{\text{trending}}$, $\mathfrak{D}_{\text{multiple}}$

(d) $\mathbf{G}_{\text{reddit}}$, $\mathfrak{D}_{10}$ (e) $\mathbf{G}_{\text{reddit}}$, $\mathfrak{D}_{101}$ (f) $\mathbf{G}_{\text{reddit}}$, $\mathfrak{D}_{\text{multiple}}$

Fig. 4. Time taken by videos to reach their final monetization state broken down by video ingestion groups ($\mathbf{G}_{\text{trending}}$ and $\mathbf{G}_{\text{reddit}}$). Red lines indicate the median and black lines indicate the 25^{th} and 75^{th} percentile.

monetization decisions being stable after 48 hours. Note, however, that the vast majority of videos experience no monetization status transitions and therefore have a stable decision upon release. Interestingly, we found that videos experiencing multiple transitions ($\mathfrak{D}_{\text{multiple}}$), on average, reached their final state faster that those experiencing only one transition (median time: 64.6 hours). This suggests that these transitions are induced by algorithmic indecision. As expected when comparing demonetized videos ($\mathfrak{D}_{10}$) and remonetized videos ($\mathfrak{D}_{101}$), we found that remonetization transitions take longer than demonetization transitions (median: 129.20 hours vs. 116.08 hours). We attribute this $\approx$ 13 hour difference to the nature of the remonetization process which involves the need for an appeal by the video creator and additional human review. Figure 4 shows the differences in the median time to final monetization state broken down by video ingestion groups. We find that, on average, videos from $\mathbf{G}_{\text{reddit}}$ appear to reach their final state faster than those from $\mathbf{G}_{\text{trending}}$. We exclude the analysis of videos in $\mathfrak{D}_{\text{multiple}}$ in our study of transition times due to ingestion source. This is because we do not believe strong conclusions can be made about time to final monetization state from the small numbers of videos in this group from each source (23 videos from $\mathbf{G}_{\text{reddit}}$ and 61 videos from $\mathbf{G}_{\text{trending}}$). For the other groups, $\mathfrak{D}_{10}$ and $\mathfrak{D}_{101}$, we found that the differences in the time to final monetization state were not statistically significant (t-test, $p > .05$).

Takeaways. While 13.3% of videos in our dataset experience non-monetization, only 0.5% experience a demonetization (monetized $\longrightarrow$ non-monetized) or remonetization (monetized $\longrightarrow$ non-monetized $\longrightarrow$ monetized) transition. Interestingly, for videos in our dataset, over one-fourth of all demonetized videos are eventual remonetized — suggesting a high false-positive rate in the demonetization algorithm. We also find that videos experiencing transitions in monetization status require, on average, five days to reach their final monetization state. In comparison to the median time for demonetization, we notice a 13-hour increase in the median time for remonetization which suggests the cost of additional human review. When considering the differences between videos ingested from Reddit-related and Trending-related sources, we find that there are statistically different rates of being monetized and experiencing demonetization transitions. However, the differences in the time to final monetization state between the groups is not statistically different. Taken together, this suggests that although videos specifically included in our dataset for their higher likelihood of being controversial ($\mathbf{G}_{\text{reddit}}$) are likely to experience different monetization decisions, the process of arriving at that final decision does not appear to be different from videos ingested specifically for their potential to have wide appeal ($\mathbf{G}_{\text{trending}}$).

5 RELATIONSHIP BETWEEN CHANNEL POPULARITY, VIDEO CONTENT, AND MONETIZATION STATUS

In this section, we focus on answering our second research question: **RQ2: How is the monetization process associated with channel and video characteristics?** Specifically, we focus on understanding how channel popularity (measured by number of subscribers) and video content are related to demonetization and remonetization rates. Taken together, our analysis brings transparency to the currently unknown influence of these factors on the demonetization algorithm.

5.1 Channel popularity

In order to understand the relationship between channel popularity and demonetization, we use the channel's subscriber count as a proxy for channel popularity and then study the rates of demonetization and remonetization conditional on this proxy. Our methods for analysis are described in §5.1.1 and our results are reported in §5.1.2.

5.1.1 Methodology. We now describe our methods for reporting demonetization rates and grouping channels by popularity.

Reporting demonetization and remonetization rates. We report the *demonetization rate* for a set of videos (V), which may all pertain to a particular topic or channel size, as the fraction of videos from the set which have ever experienced a demonetization transition:

$$\text{DemonetizationRate}(V) = p(\text{demonet.} \mid V) = \frac{|(\mathfrak{D}_{10} \cup \mathfrak{D}_{101}) \cap V|}{|V|}. \tag{1}$$

We quantify the *remonetization rate* as the likelihood that a demonetized video in V will be remonetized:

$$\text{RemonetizationRate}(V) = p(\text{remonet.} \mid \text{demonet.}, V) = \frac{|\mathfrak{D}_{101} \cap V|}{|(\mathfrak{D}_{10} \cup \mathfrak{D}_{101}) \cap V|}. \tag{2}$$

We choose the this approach because it captures the notion of an 'error-admission' rate – i.e., each time a demonetized video is remonetized, it is indicative of an admission of incorrect determination made by the demonetization algorithm. By leveraging this definition of a remonetization rate, we are able to analyze the video content that is likely to trigger incorrect algorithmic decisions and the channels for which YouTube is more likely to change their algorithmic decisions.

Grouping demonetized and remonetized channels by popularity. In order to analyze the relationship between channel popularity and the characteristics of monetization status transitions, we group channels by their subscriber size. More specifically, we bin channels which experience monetization status transitions into three categories based on their subscriber counts: those with between 1K and 100K subscribers (P_1), those with between 100K and 1M subscribers (P_2), and those with over 1M subscribers (P_3). Table 4 shows the breakdown of monetization status transitions observed by channel popularity group and video ingestion source. Due to the small number of $\mathfrak{D}_{\text{multiple}}$ transitions observed in each popularity group and the subsequent lack of significance in our results, we do not report analysis on videos in this group.

5.1.2 Results. We now present our results on the relationship between popularity and rates of demonetization and time to final monetization state. We note that we do not break down our analysis by the video ingestion source ($\mathbf{G}_{\text{trending}}$ and $\mathbf{G}_{\text{reddit}}$) because of the small numbers that arise in each popularity category once we do so (e.g., from Table 4, we can see that $\mathbf{G}_{\text{reddit}}$ has only

Ingestion group	Popularity group	# Channels	$\mathfrak{D}_{10}$ events	$\mathfrak{D}_{101}$ events	$\mathfrak{D}_{multiple}$ events
$\mathbf{G}_{trending}$	Combined	326	725	400	59
$\mathbf{G}_{reddit}$	Combined	139	332	68	23
Combined	P_1: [1K, 100K)	119	277	69	9
	P_2: [100K, 1M)	192	549	237	30
	P_3: [1M, ∞)	152	231	162	43
$\mathbf{G}_{trending}$	P_1: [1K, 100K)	63	142	47	7
	P_2: [100K, 1M)	139	449	209	21
	P_3: [1M, ∞)	124	134	144	31
$\mathbf{G}_{reddit}$	P_1: [1K, 100K)	56	135	22	2
	P_2: [100K, 1M)	54	100	28	9
	P_3: [1M, ∞)	29	97	18	12

Table 4. Breakdown of monetization status transitions observed by channel subscriber counts and ingestion source. We only consider channels which experienced a monetization status transition.

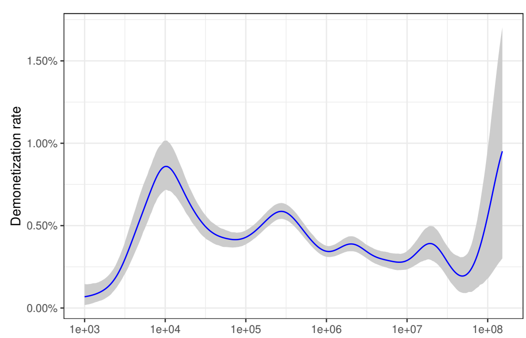


Fig. 5. Demonetization rate by channel popularity (measured by number of subscribers).

18 $\mathfrak{D}_{101}$ events and 29 channels in P_3). We do not believe it would be responsible to make inferences or comparisons about the demonetization process based on our ingestion sources as a result.

Relationship between popularity and de-/re-monetization rates. Table 5 reports the demonetization and remonetization rates broken down by channel sizes (measured by number of subscribers). From these results, we observe a clear trend: larger channels belonging to the P_3 popularity group (> 1M subscribers) appear to have lower rates of demonetization and higher rates of remonetization than smaller channels. These differences were found to be statistically significant when compared with channels in P_1 and P_2 (two proportions z-test, p<.05). These differences become apparent in Figure 5 which shows how the demonetization rate changes as channel popularity increases. We see that, with the exception of the tails (channels with under 10K subscribers and over 10M subscribers), the rate of demonetization appears to decrease as popularity increases. We note that the large increase in demonetization rate for channels with over 10M subscribers is caused by a large number of demonetization events attributed to a single channel — PewDiePie, a creator known in part for their controversial content [1, 91].

Relationship between popularity and time taken for transition to final monetization status. Figure 6 shows the time taken for videos uploaded by channels of different sizes (measured

Ingestion group	Popularity group	Demonetization rate	Remonetization rate
$\mathbf{G}_{\text{trending}}$	Combined	0.45%*	35.6%*
$\mathbf{G}_{\text{reddit}}$	Combined	0.39%*	17.0%*
Combined	P_1: [1K, 100K)	0.47%	19.9%*
	P_2: [100K, 1M)	0.49%	30.2%*
	P_3: [1M, ∞)	0.33%*	41.2%*

Table 5. Demonetization and remonetization rates broken down by channel subscriber counts and ingestion source. We do not report demonetization and remonetization rates for the *intersection* of subscriber counts and ingestion sources due to the very limited number of $\mathcal{D}_{101}$ events (see Table 4). Here * indicates p-value< 0.05 for a two proportions z-test. We ran this test on all pairwise demonetization and remonetization rates within groups (e.g. $\mathbf{G}_{\text{trending}}$ vs. $\mathbf{G}_{\text{reddit}}$ and P_2 vs. P_3), and all were statistically significant except P_1 vs. P_2.

by number of subscribers) to transition to their final monetization state. The differential treatment of smaller channels becomes immediately apparent. Figure 6a- 6c show that smaller channels (P_1) experience demonetization transitions faster than larger channels (P_1 median: 65 hours, P_2 median: 134 hours, P_3 median: 109 hours). However, these differences were not fouund to be statistically significant. Conversely, Figure 6d- 6f show that the time for remonetization transitions is much longer for smaller channels (P_1 median: 228 hours, P_2 median: 133 hours, P_3 median: 113 hours). These differences were found to be statistically significant when comparing for all pairs of popularity groups (t-test, p-value < .05). Taken together, these results show that smaller channels remain non-monetized longer than larger channels, regardless of the "correctness" of the algorithmic demonetization decision. Taken together, our results lend credibility to creators' concerns about the preferential treatment to larger channels: smaller channels are more likely to experience higher rates of demonetization through a speedier process while also experiencing lower rates of remonetization through a slower process.

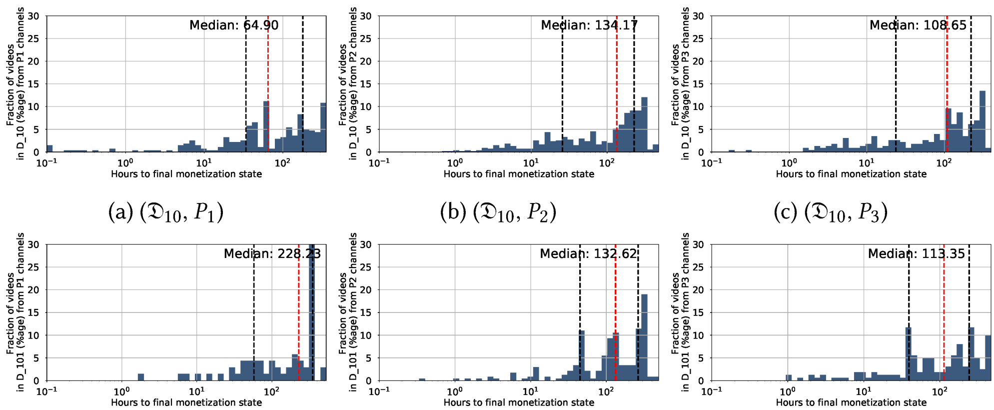
Takeaways. The results of our analysis suggests that the monetization process is different for channels of different subscriber counts. Specifically, we find that channels with smaller subscriber counts experience higher rates of demonetization and longer times to remonetization than larger channels when compared with popular channels having more than 1M subscribers. These findings support existing theories of the presence of a tiered governance structure on YouTube [32].

5.2 Video content

We now focus on understanding the role of video content on demonetization. In order to do so, we study the rates of demonetization and remonetization conditional on word-type occurrences and video topics obtained from video titles and video captions. We use video titles and video captions because, after manual inspection, they were found to accurately capture the content of a video. We made the decision to not rely on video descriptions because manual inspection found them to not accurately reflect the content of the video. In particular, we found that they often contained large numbers of 'tags' unrelated to the video (perhaps for reasons related to search engine optimization). Our methods for analysis are described in §5.2.1 and our results are reported in §5.2.2. We use the same methodology described in §5.1.1 to report demonetization and remonetization rates.

5.2.1 Methodology. Our analysis of video content is conducted by analysis of the video titles and captions as described below.

Conditioning on word-type occurrences in video titles and captions. In order to analyze the relationship between video content and demonetization, we begin with the following simple



(a) $(\mathfrak{D}_{10}, P_1)$ (b) $(\mathfrak{D}_{10}, P_2)$ (c) $(\mathfrak{D}_{10}, P_3)$

(d) $(\mathfrak{D}_{101}, P_1)$ (e) $(\mathfrak{D}_{101}, P_2)$ (f) $(\mathfrak{D}_{101}, P_3)$

Fig. 6. Time taken by videos to reach their final monetization state broken down by channel subscriber counts. Red lines indicate the median and black lines indicate the 25^{th} and 75^{th} percentiles.

question: *Do unigram word-types in video titles or captions have an association with observed demonetization and remonetization rates?* The results of such an analysis will help validate current theories that YouTube maintains a "blacklist" of words whose presence in content result in demonetization or non-monetization. For this analysis, we compute the probability of the demonetization and remonetization rates of the set of videos that contain at least one instance of a specific word w. Or equivalently, it is calculated over videos (i) conditional a specific word-type (w) occurring in the video's content (i^{content})—either its title or captions,

$$Pr(D \mid w) = \frac{\sum_{i \in D} 1\{w \in i^{\text{content}}\}}{\sum_{j \in V} 1\{w \in j^{\text{content}}\}} \tag{3}$$

where $D \in \{\mathfrak{D}_{101}, \mathfrak{D}_{10} \cup \mathfrak{D}_{101}\}$ and V is the set of all videos in our dataset. We only consider words which occur in at least 50 of the videos in our dataset (7,878 words). We tokenize titles and captions with *NLTK* [64].

In addition to conducting this analysis on individual word-types, we also analyze video rates for three pre-defined categories speculated to have high demonetization rates: Black Lives Matter (BLM) protests, the COVID-19 pandemic, and LGBTQ issues. We selected these topics because of the public controversies and allegations of bias against their related content (e.g., [3, 85]). To create these categories, subject-matter experts including journalists and media scholars created lists of seed keywords for each category. Then we train GloVe word embeddings [76] on all the videos' titles and captions in our dataset. Using these trained embeddings, we infer a embedding for each seed keyword and find its top 20 nearest neighbors by cosine distance. The same domain experts examine the nearest neighbor reports and manually select additional valid keywords. Table 6 shows the final keywords for the three topics. We calculate the probabilities analogously to Equation 3, but count if the video's title or caption matches *any* word in the category. Links to videos whose titles or captions were identified as having a match with any of the final keywords were independently sampled by the authors of this paper to confirm that their content reflected the matching keywords.

Conditioning on video topics from captions. In order to assess the relationship between video topic and monetization decisions, we train a topic model [30] on all the videos with English

Category: Keywords	**# Videos**	**Demonetization rate**	**Remonetization rate**
BLM: blm floyd brutality kenosha acab antifa anti-fascist antifascist protest protests protester riot riots looting	4,182	0.496%	20.1%
Covid-19: corona covid covid-19 covid19 coronavirus rona kovid chronovirus wuhan lockdown	10,111	0.389%	33.1%
LGBTQ: lgbtqa lgbtqia lgbtq+ lgbtqa+ lgbtqia+ homophobic homo homosexual homosexuals homosexuality homophobia lgbt bisexual bisexuals bisexuality asexual intersex transsexual pansexual cisgender genderqueer transgender transphobic gay gays cis bi lesbian lesbians trans transphobia polyamory polyamorous non-binary gender-fluid queer	740	0.512%	21.2%

Table 6. Manually curated list of keywords surrounding Black Lives Matter (BLM), Covid-19, and LGBTQ issues and the demonetization and remonetization rates associated with video captions containing *any* of the associated category keywords. We also report the total number of videos across our entire dataset whose captions matched any of the keywords (# Videos).

captions in our dataset.[2] We leverage captions because they broadly represent a video's content, unlike the title or description of a video which content creators may craft to entice viewers and may not necessarily be representative of the video's full content. Topic models are statistical models that can learn latent clusters of words from pieces of text, and topic models have been used extensively for analyzing other content such as disinformation on social media [88], scientific articles [46], and political texts [47, 82]. We apply standard pre-processing from natural language processing to video captions: we tokenize with *Gensim* [77], discard any tokens that are less than 2 characters or greater than 25 characters (to remove noisy text from auto-generated captions), remove punctuation, lowercase, use *NLTK* [64] to remove stop words, and remove tokens that occur in fewer than 10 videos or in the top 5% of videos. This results in a vocabulary of size 100,957. We then use *MALLET* [66] to train a model with 300 topics. Once trained, our model infers a multinomial distribution over the 300 topics, θ_i, for each video i. We then calculate the probability of a group $D \in \{\mathfrak{D}_{11}, \mathfrak{D}_{00}, \mathfrak{D}_{10}, \mathfrak{D}_{101}\}$ conditional on topic, k:

$$p(D \mid \text{topic } k) = \frac{\sum_{i \in D} \theta_{i,k}}{\sum_{j \in V} \theta_{j,k}} \tag{4}$$

This can be used to calculate demonetization and remonetization rates conditional on a particular dimension of videos' topical mixed memberships, directly comparable to the video subset-conditional rates from Equations (1) and (2). The authors of this paper performed manual inspection over a sample of the 300 topics and the set of videos (from each monetization status type) which were found to have the five highest θ's for the corresponding topics. The topics were found to satisfactorily reflect the content of the video. Video descriptions were excluded from our analysis because of their poor performance during this manual inspection.

5.2.2 Results. We now present our analysis of the demonetization rates based on word-type occurrences and video topics. For the same reasons as §5.1, we do not include a breakdown of results by ingestion source.

[2]We limit our analysis to the 211,797 videos (68% of all videos in our dataset) that have either manual or auto-generated English captions.

Word	#Videos	Demonetization rate
spiderman	56	30.3%
mobs	67	29.8%
mcgregor	118	25.4%
euphoria	50	24.0%
conor	119	21.8%

Table 7. Word-types in video titles with the highest demonetization rates. Remonetization rates were observed to be 0% for all listed word-types.

Demonetization and remonetization rates conditional on word-types. Table 7 and Table 6 show the demonetization and remonetization rates of videos containing specific word types in their captions and titles respectively. Although we find instances of word-types occurring with high demonetization rates (Table 7), the fact that there is no single keyword resulting in a 100% demonetization rate suggests that demonetization decisions are not made solely based on the presence of specific words in titles. Examining the top-ranked demonetized words in Table 7, the occurrence of 'spiderman' as our word-type with the highest demonetization rate is notable since it has been reported in recent work [73] that Spiderman frequently made an appearance in inappropriate videos targeted at children. The other popular word-types were found to be related to very popular sporting events or films, suggesting copyright infringements as the reason for their high demonetization rate. Applying the word-type unigram analysis to captions had similar results.[3] Table 6 shows the demonetization rates for our three selected categories and their associated keyword lists applied to videos' captions.[4] We see that the demonetization rates for the BLM category (0.496%) and the LGBTQ category (0.512%) are roughly equivalent to the baseline demonetization rate (0.5%), but the remonetization rates for BLM (20.1%) and LGBTQ (21.2%) are lower than the baseline (30.1%). The COVID-19 category is the opposite with a demonetization rate (0.389%) lower than the baseline, but a remonetization rate (33.1%) slightly higher. Although we see no evidence of alarming rate differentials with these topics, we note that our data was gathered at a different time from when the allegations of bias and differential treatment pertinent to these topics were made.

Demonetization and remonetization rates conditional on video topics. Figure 7 and Table 8 show the results for topic-modeling applied to video captions. In Figure 7, we see there are several outliers with much higher demonetization rates per topic such as *Topic F* and *Topic D* which correspond to Islam and basketball topics, respectively (top words associated with each topic are shown in Table 8). Notably for these topics the remonetization rate is low. Upon manual inspection of the videos most reflective of the topic, we were unclear as to the reason for demonetization of videos matching the Islam topic (*Topic F*). However, we found that videos matching the basketball topic (*Topic D*) were screen-captures and highlights of NBA games – likely demonetized due to copyright infringements (note that our videos were gathered during the 2020 NBA Playoffs season). Moving our attention to the top-left quadrant in Figure 7, we see that *Topic A* and *Topic B*, which roughly correspond to wrestling and Super Mario games have high demonetization rates and high remonetization rates. In our manual inspection, we found that videos related to *Topic A* that were remonetized were centered around professional wrestling with creators sharing personal opinions

[3] We omit for brevity. Topic models applied to captions gave similar results.

[4] We also did the category analysis for videos' titles, however there were too few matches for significant results. For $\mathfrak{D}_{101}$ there were only 2 videos whose titles matched keywords in the BLM cateogry, 5 for COVID, and 2 for LGBTQ. For $\mathfrak{D}_{10}$ there were only 26 videos whose titles whose titles matched keywords in the BLM category, 15 videos in the COVID category, and 2 in the LGBTQ category.

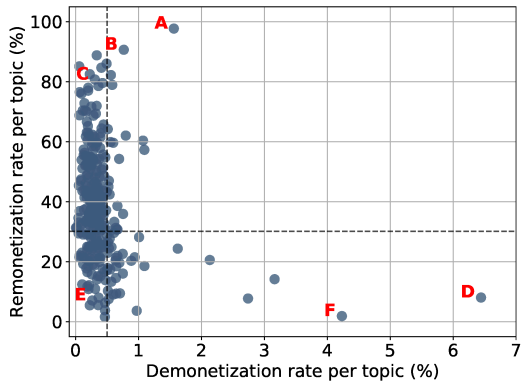


Fig. 7. Per-topic demonetization rate and per-topic remonetization rate. Each circle on the plot is one of 300 topics inferred from the English captions of videos. Red topic labels correspond to the selected topics in Table 8. Dashed lines indicate the mean per-topic demonetization and remonetization rates.

Topic k	**Top 10 words**	**Demonetization rate**	**Remonetization rate**
A	wwe wrestling raw nxt vince summerslam champion smackdown sasha seth	1.56%	97.7%
B	mario nintendo kart rc coins treasure luigi toad donkey hammer	0.77%	90.6%
C	spider marvel hulk avengers iron captain tony thor stark shield	0.31%	80.7%
D	lebron lakers anthony bubble clippers skip davis playoffs portland lillard	6.44%	8.1%
E	biden democratic democrats bernie republican voters republicans voting sanders candidate	0.26%	7.1%
F	muslim allah islam muslims quran al prophet muhammad islamic religion	4.23%	1.9%
	Baseline Rates	0.50%	30.1%

Table 8. Selected topics (corresponding to the red callouts in Figure 7) along with the inferred top 10 words per topic. *Baseline rates* are the average across all topics weighted by the number of videos in that particular transition group.

around professional wrestling events and news while demonetized videos appeared to be showing copyrighted WWE content. We were unable to explain the demonetization and remonetization rates of the Super Mario topic (*Topic B*). Other categories associated with high remonetization rates can be found in the bottom-left quadrant. We selected *Topic C* associated with the Marvel superhero franchise for manual inspection of remonetized videos. We were unable to understand the reasons that might have caused an initial demonetization decision. We also highlight political videos (seen in *Topic E*) had a low demonetization and remonetization rate – a finding that supports the absence of political bias in YouTube's platform moderation also seen in other studies [54, 55].

Takeaways. The results of our keyword (word-type) analysis suggests that demonetization decisions are not based on the occurrence of specific keywords in video titles. Our topic analysis indicates that certain topics are subject to a much higher rate of demonetization than others. Manual inspection shows that some of these demonetization are justified (e.g., copyright infringements). However, other decisions are not immediately explainable (e.g., Topics B, C, and F). We also find evidence of the existence of a potentially problematic bias with videos related to the Islamic religion having very high rates of demonetization and low rates of remonetization.

6 RELATIONSHIP BETWEEN DEMONETIZATION AND CHANNEL GROWTH

In addition to analyzing the properties of channels and videos affected by demonetization and remonetization decisions, we also examine: **RQ3. How is demonetization associated with a channel's future growth rate?** We hypothesize that an initial demonetization transition could reduce a video's visibility and the associated channel's views – representing a harm to creators and their revenue.

6.1 Methodology

Measuring channel views and growth rate. When a new channel is added to our dataset from any of the avenues described in §3.1, our system continually downloads its *subscriber count* and overall *view count*, which are displayed on the channel's home page, derived from the sum of all views for currently published videos.[5] All channels typically see growth in *view count* over time, including over the several months of time in our dataset. Since we only have observational data, estimating the causal effect of demonetization is difficult since many latent confounding variables may affect both the probability of demonetization as well as the outcome of popularity. Still, we find it an instructive, if preliminary, step to assess how channel view growth changes before and after a demonetization, compared to growth changes for videos that do not experience a demonetization.

Let the outcome, channel view count over time, be $y_{c,t}$ for channel c at time t. Let $t_i^{(0)}$ be the time of demonetization of video i. We calculate the *change in growth rate* for i as the ratio of pre- and post-demonetization growth rates over 7-day periods before and after the demonetization event:

$$R_i = \frac{\text{Post-Event Growth}_i}{\text{Pre-Event Growth}_i} = \frac{y_{c,t_i^{(0)}+7} - y_{c,t_i^{(0)}}}{y_{c,t_i^{(0)}} - y_{c,t_i^{(0)}-7}} \tag{5}$$

By itself, $R_i < 1$ indicates the descriptive fact of growth slowing around the event.

Obtaining control and treatment groups. We compare the growth changes of channels that experience a video demonetization event to a reasonable baseline of channels that do not experience a video demonetization event. More concretely, the "treatment" group, $\mathcal{T}$, consists of the set of videos that are demonetized ($\mathfrak{D}_{10} \cup \mathfrak{D}_{101}$). The "control" group, C, consists of a subset of always-monetized videos ($\mathfrak{D}_{11}$), whose channels never experienced a video demonetization event in our dataset. For each control video i, we define a synthetic counterfactual demonetization time, $t_i^{(0)}$, as 7 days after its addition time to our dataset, representing when it could have been demonetized. For videos in $\mathcal{T}$, the average time from the added date to $t_i^{(0)}$ is 4.9 days, so 7 days is a reasonable time to expect to see a demonetization event occur while controlling for day-of-week effects on view count. Comparing the growth rates of $\mathcal{T}$ and C helps us get closer to answering our ideal experimental question, *"What would have happened to channel views if a demonetized video was not demonetized?"*, corresponding to the *treatment effect on the treated* in the causal inference literature

[5]Unfortunately, due to a technical limitation we were unable to collect views at the video level.

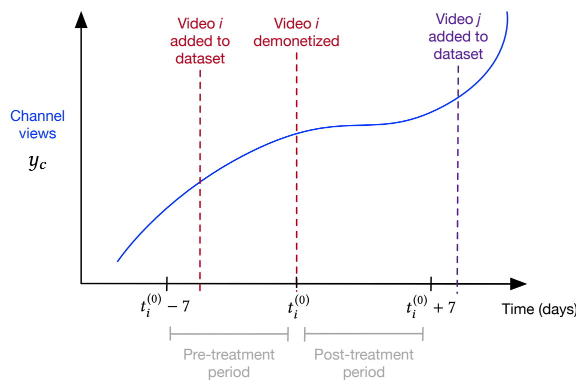


Fig. 8. Example view trend for a single channel. We aim to measure a channel's view growth rate in the *pre-treatment period*—before the target video, video i, is added to our dataset and prior to its demonetization–and compare it to the view growth rate in the *post-treatment period*—after the video has been algorithmically demonetized ($t_i^{(0)}$). In our analysis, we simplify the pre-treatment and post-treatment periods to be $t_i^{(0)} \pm 7$ days. The same channel may have other videos, video j for example, added before, during or after the time period of interest.

[69]. Figure 8 illustrates one (contrived) example of a channel's view count over time, new video uploads to the channel, and a demonetization event.

Overall comparison of growth change conditional means. To analyze the video groups, we report the average treatment and control growth changes, and their difference,

$$E[R_i \mid i \in \mathcal{T}], \quad E[R_i \mid i \in C], \quad \Delta^{\text{diff}} = E[R_i \mid i \in \mathcal{T}] - E[R_i \mid i \in C]. \tag{6}$$

Ingestion source and channel popularity breakdown. As shown in §5.1, the popularity of a channels and the source from which a channel and its videos were ingested into our dataset (i.e., controversial subreddits or popular trending videos) affect the rate of demonetization. Therefore, they may also be confounders in our analysis. After all, the dynamics of view growth may vary: (1) across different popularity groups (P_1, P_2, P_3) or (2) across our ingestion mechanisms ($\mathbf{G}_{\text{reddit}}$ and $\mathbf{G}_{\text{trending}}$) which effectively serve as a proxy for types of content in videos and separate videos that may be affected by YouTube's policies for "trending" channels. Thus, we estimate the effects of a video being demonetized within: (1) groups of channels with similar subscriber counts and (2) groups of channels from the same ingestion source. We report the differences of conditional means within these groups, e.g., $\Delta^{\text{diff}}_{P_1} = E[R_i \mid i \in \mathcal{T} \cap P_1] - E[R_i \mid i \in C \cap P_1]$. We note that we do not further break down the groups of channels from the same ingestion source by their popularity because of the small number of demonetization events that would arise for each subgroup (Table 4).

Population selection. We include several other filters to select the population of channels that are valid to use in our analysis: we filter to channels for which we have at least fourteen days of view measurements, and to videos whose channels have ±7 days of view data around the video's $t^{(0)}$. We also remove channels whose view count time series decrease at any point, presumably due to the channel removing videos (approximately 25% of all channels). We also remove channels with greater than 500 total videos, since these channels are too large to infer the influence of a single video being demonetized. This results in 226 demonetized videos comprising $\mathcal{T}$ (from 109 unique channels), and 77,166 always-monetized videos for C (from 4412 unique channels).

Confidence intervals and statistical significance. We calculate bootstrap confidence intervals [37, 41] by resampling both treated and control videos with replacement, using the percentile method to form a 95% confidence interval with the 2.5% and 97.5% bootstrap percentiles. We use 10k bootstrap samples. We also calculate p-values for a two-sided Welch's t-test to test for a null difference.

Limitations. Like any attempt to infer causal effects from observed data, validity is threatened by unobserved confounders that could affect both treatment (video demonetization) and outcome (a channel's popularity dynamics) [49, 69, 75]. One potential confounder is other videos being uploaded to the channel around the same time of the target (demonetized) video, thus making it difficult to isolate the effect of a single video. However, we find this confounder is roughly balanced between treatment and control groups: for a single channel, the median number of videos added per day for $\mathcal{T}$ is 0.89 and for C is 0.77. Another potential source of confounding are properties of the target video that could influence an algorithmic demonetization decision, such as their content. Numerous other studies have used text to remove latent confounding for causal estimates; see Keith et al. for a review [58]. We attempted to condition on video content using coarsened exact matching on the topic model representation of videos from §5.2, as has been done for in an analogous setting, internet censorship analysis [81]. However, caption data, the prerequisite for topic modeling, were available for only 88 of the 226 treated videos. These 88 were unrepresentative of the broader population, having a significantly different distribution of growth changes ($E[R|\mathcal{T}] = 1.24$; compare to $E[R|\mathcal{T}] = 0.90$ in Table 9's first row). Since we are currently uncertain how caption availability interacts with popularity or other factors on YouTube overall, we leave this analysis for future work.

6.2 Results

Relationship between demonetization and channel growth rate. Table 9 shows the growth change for the set of demonetized videos ($\mathcal{T}$), monetized controls (C), and the estimated effects (Δ^{diff}). On average, channels see a −9.9% decline in views after a video is demonetized ($E[R|\mathcal{T}] - 1$), compared to a baseline of slightly positive growth, 1.9%, among controls. This results in an average effect of −11.8 percentage points. Since the treatment set is not large, the confidence interval for the difference is wide, but statistically significant (CI=[−21.5, −0.9], $p = 0.03$). Our analysis broadly suggests that channels experiencing demonetization events do in fact experience a decline in view growth rate (a negative treatment effect) than those that do not — indicating a potential harm to channels. When we examine the breakdown by ingestion source, we see that both $\mathbf{G}_{\text{trending}}$ and $\mathbf{G}_{\text{reddit}}$ have relatively similar average effects to each other (−13.2 and −9.5 percentage points respectively) and to the overall dataset, although these confidence intervals are very wide. This seems to indicate the ingestion source of the channel does not have a large influence on the effects of demonetization on videos. When we examine the breakdown by channel popularity (P_1, P_2, P_3), the only statistically significant effect is for P_2 which has a large negative average effect of −30.1 percentage points.

Takeaways. Our analysis shows a statistically significant difference in channel view growth rate based on monetization status. This difference is statistically significant which lends additional credence to the hypothesis that demonetization of a video results in fewer recommendations proposed in prior work by Kumar [61] and Caplan & Gillespie [32].

Dataset	Growth change ($E[R]-1$)		Avg. effect (Δ^{diff}), 95% CI	
	Demonet. ($\mathcal{T}$)	Monet. (C)		
All data	−9.9%	+1.9%	−11.8 pp	[−21.5, −0.9]*
Within $\mathbf{G}_{\text{trending}}$	−13.0%	−0.2%	−13.2 pp	[−25.7, +2.1]
Within $\mathbf{G}_{\text{reddit}}$	−1.9%	+7.6%	−9.5 pp	[−20.8, +2.8]
Within P_1: [1K, 100K)	+21.6%	+5.5%	+16.1 pp	[−1.4, +36.0]
Within P_2: [100K, 1M)	−29.9%	+0.2%	−30.1 pp	[−43.7, −12.1]*
Within P_3: [1M, ∞)	+7.8%	+2.4%	+5.4 pp	[−4.6, +16.3]

Table 9. Average growth rate change for a channel, before and after the demonetization of one video, and the difference in growth change (Δ^{diff}) that could be attributed to demonetization. Growth change is the average ratio $E[R]$ minus 1, displayed as a percentage; thus the effect size is in percentage points (pp). For all videos, $\mathcal{T}$ consists of 226 videos from 109 channels and C consists of 77,166 videos from 4,412 channels. In the size breakdown, $\mathcal{T}$ is 56 videos from 36 channels for P_1; 126 videos from 44 channels for P_2; and 44 videos from 29 channels for P_3. In the ingestion source breakdown, $\mathcal{T}$ is 163 videos from 74 channels for $\mathbf{G}_{\text{trending}}$; and 63 videos from 35 channels for $\mathbf{G}_{\text{reddit}}$. For all of these supgroups, C is greater than 13,000 videos from over 1,000 channels. Here, * indicates p-value<0.05 for Welch two-sample t-test between the treatment and control rates for a particular dataset.

7 RELATED WORK

Automated text analysis and causal inference methods for studying online systems. In studying videos' content, this work aligns with methods from natural language processing [43, 56] and automated analysis of text as social data [47, 70, 72]. Methods similar to the ones we present in Section 5.2 have been used to examine content on other online platforms such as noisy language on Twitter [42], hate speech on Reddit [33], and troll-like comments on online news sites [36].

Although we acknowledge we have not accounted for all unmeasured confounding necessary to claim our analysis in Section 6 as causal, our analysis is inspired by work that attempts to infer causal effects from observational data [49, 69, 75]. Our analysis of growth rates before and after a demonetization event is inspired by *difference-in-differences* approaches [21, 26]. Other researchers have also attempted to measure causal effects from observational data on online platforms, such as estimating the effect of alcohol use on academic performance via users' Twitter posts [60]. Youtube demonetization is structurally similar to post-publication moderation by a platform, which is typically applied to delete posts the platform finds objectionable; in this setting, a number of works have studied Chinese internet censorship, including what topics tend to be deleted [28, 65, 80], and Roberts et al. develop a textual causal inference method, topical coarsened exact matching, to assess effects of censorship on users [81]. For Reddit, Chandrasekharan et al. use difference-in-differences methods to estimate the effect of Reddit banning certain communities on the users' future hate speech volume [33].

Computational audits of YouTube's algorithms. Although there have not been any prior computational audits of YouTube's demonetization algorithms, there have been efforts to audit related automated processes such as comment moderation, search, and content recommendation. Most closely related to our work are the efforts of Yin and Sankin [93] to understand the interfaces provided to advertisers, by YouTube, for aiding selection of videos on which ads would be displayed. They found evidence suggesting the use of blocklists to prevent placement of ads on problematic videos. Our study differs in that we approach the study of demonetization from the perspective of content creators rather than advertisers. Although our analysis did not find evidence to support a keyword-based filter for demonetization decisions, it should be noted that in addition to leveraging a different measurement perspective our analysis was conducted on data gathered prior to the study by Yin and Sankin. In a recent series of studies, Jiang et al. [53–55] focus on comment moderation processes of YouTube and find that claims of political biases in the moderation process are unfounded. Instead, they find that algorithmic moderation efforts are predicted by the presence of misinformation, hate speech, and other forms of extremism. Several studies have also sought to audit the recommendation algorithms leveraged by YouTube. Focusing on the appropriateness of content recommended to children, Papadamou et al. [73] developed a classifier to identify inappropriate content targeted at children. Using this classifier, they showed that the likelihood of a toddler being recommended an age-inappropriate video after watching benign, but related, content was high (between 1.3% to 3.5%). In similarly related work, Araujo et al. [27] studied the content, advertisement, and audience characteristics of several large YouTube channels producing content targeted at children. Using AI tools to identify demographics and gender from public audience profiles, they found that a large number of the profiles leaving public comments on these videos were under the age of 12. Their findings highlighted potential violations of regulations aimed at protecting the privacy of and prohibiting advertising targeted at children such as COPPA (Children's Online Privacy Protection Act) in the United States and the Consumer Code in Brazil. Hussein et al. [51], focused on the recommendation of videos containing misinformation, show that the recommendation algorithms leveraged by YouTube appear to be insensitive to biases due to user demographics. However, they found that these algorithms exhibit a filter bubble effect wherein a users watch history has a high similarity to the subsequently recommended videos – even in the case of videos containing conspiracies and misinformation. Similar studies have focused on measuring the recommendation algorithm's tendency to promote other types of problematic content such as alt-right extremist content [78], incel content [74], and clickbait [98].

Social science research on the impact of YouTube's demonetization algorithms. The impact of YouTube's monetization algorithms and policies have been studied in the social sciences from two perspectives: (1) impact on creators and (2) impact on social and cultural trends. Both perspectives highlight the reduced stability for creators and the increased tendency towards (self-) censorship as consequences of the opaque demonetization process.

Impact on creator participation and attitudes. Caplan and Gillespie [32] explored YouTube's monetization policies as a form of tiered governance in which established media partners and amateur creators experience different treatment and policies. Leveraging a dataset of 90 videos where creators describe their challenges with demonetization and their own interviews with several YPP-members, their work highlights how information asymmetries and inconsistencies present in the demonetization process lead to beliefs of being algorithmically controlled or censored and an emotional toll on creators. A similar approach is used by Kumar [61] to investigate the tensions between creators and YouTube's algorithms and the impact of demonetization on channel revenue and visibility. Importantly, this work hints at a relationship between the demonetization and recommendation algorithms leveraged by YouTube – i.e., demonetized content appears to be less likely to be recommended. Our study which performs a large-scale audit of YouTube's demonetization algorithms shows that the monetization experiences of smaller channels are indeed different than those of larger channels. Specifically, we find that smaller channels experience more demonetization at a faster rate (§5.1) and demonetization does appear to result in a significant reduction in channel view growth rate (§6). These findings lend additional support to the analysis and conclusions of Caplan & Gillespie [32] and Kumar [61]. On a smaller scale, Sehl and Ross [84] performed a case-study on *The Philip DeFranco Show*, a popular independent YouTube news channel with over 6M subscribers. Their research highlights how the actions and behavior of content creators change (particularly self-censorship and changing video release patterns) as a result of the algorithm-driven demonetization process. These findings were reinforced by a qualitative study by Stanford [87] which found that in addition to the change in creator habits and dynamics between the platform and creators, the demonetization algorithm also resulted in different types of engagement and increased openness about revenue and content monetization between creators and their audiences.

Contribution to shifts in social and cultural trends. Numerous studies [25, 59, 68] have also considered the power imbalances present on digital platforms, with YouTube's monetization process as an example, as a means for commercialization and exploitation of labor. The general focus of these studies is on the imbalances presented by the fact that participation is the only decision made by the creator while all other decisions and information related to commodification and monetization of content are left in the hands of the platform. Despite the existence of such unfavorable asymmetries facing content creators, studies have also highlighted the popularity of the YouTube platform and its significant impact on the news and media industry by contributing to the "gig-ifying" and platformization of cultural production and lower creator stability [24, 45, 71]. Demonetization has also been explored as an extension of corporate censorship. Wilkinson and Berry [90] used a media-studies framework to study YouTube's demonetization of LGBTQ-related content as a form of censorship by proxy. Their work showed that the predictors for and effects of support of such censorship are different from censorship by the state. Along similar lines, Alkhatib [22, 23] leveraged anthropological theories of bureaucracies to explore the tendency for algorithmic governance systems, with YouTube's monetization framework as an example, to censor and disempower the already marginalized in society. Put in the context of the above work, our analysis on the occurrence of word-types and topics in demonetized and remonetized videos (§5.2) did not find that videos associated with marginalized groups (specifically, BLM and LGBTQ+) were disproportionately impacted by YouTube's demonetization algorithms. This finding does not suggest the use of demonetization as a form of censorship for these specific groups. However, it is important to note that this conclusion is drawn from a small number of videos (4,182 BLM-related and 740 LGBTQ-related videos) which were gathered from a specific six-week period.

8 DISCUSSION

Limitations. Fundamentally, this work is a 'best-effort' large-scale measurement study aimed at bringing transparency to YouTube's monetization algorithms. Our study comes with limitations that hinder the accuracy of some of our inferences and its subsequent findings. As explained in §3.2, our monetization status inference technique: (1) has the potential to miss interesting transitions in monetization status; (2) is unable to distinguish creator-driven monetization decisions from algorithm-driven monetization decisions; and (3) can only guarantee checks on the monetization status of videos once every two hours. Although we take steps to mitigate the effects of such errors by restricting our analysis only to videos which experienced an observed and lasting demonetization or remonetization transition (i.e., $\mathfrak{D}_{10}$ and $\mathfrak{D}_{101}$), we are ultimately restricted in our ability to completely understand the algorithm. For example, our analysis methods could not be applied to: (1) $\mathfrak{D}_{00}$ and $\mathfrak{D}_{01}$, the largest datasets of non-monetized and transition group videos, due to the possibility of errors, or (2) $\mathfrak{D}_{\text{multiple}}$ due to its small size, or (3) consider the impact of ingestion sources *and* popularity groups simultaneously due to the small number of videos resulting in each subgroup. Our relationship analysis approaches are also limited due to possible confounders that arise from the high rate of videos released by the channels in our dataset, making it hard to isolate the impact of a single demonetization event on the channel views growth rate. Further, our analysis on the relationship between video content and demonetization was exclusively text-based (video titles and captions). Prior work [98] has suggested that non-text metadata such as video thumbnails may influence the characterization of a video and it remains unclear if this influence extends to the demonetization process. Future work may consider the inclusion features drawn from these thumbnails upon which demonetization is conditioned. Finally, our data was collected during a six-week period from July 9th to September 22nd, 2020. Although there were no changes in YouTube's monetization policies during this time, the mechanisms for enforcement of their policies underwent drastic changes. Specifically, it was announced that YouTube relied more heavily on AI-based moderation tools due to the unavailability of human reviewers during the ongoing Coronavirus pandemic [57, 94]. This introduces several complicating factors in our study. For example, we have no way of identifying which videos were the subject of additional human review, or whether the observed incidence rates and time to final state of $\mathfrak{D}_{10}$, $\mathfrak{D}_{101}$, and $\mathfrak{D}_{\text{multiple}}$ monetization transitions are different than when human reviewers are more involved in the process. Despite these limitations, we argue that there is value in our measurements since they provide rare insight into the decisions made by a largely AI-driven monetization process that impacts millions of creators and users.

Regulating online platform moderation. Calls to regulate moderation on online platforms have emerged from both sides of the political spectrum in the US. As has been shown in prior work, demonetization results in self-censorship by creators – in effect making it a tool for moderation and control on a platform. This provides platforms which share revenue with contributors with two options for action: outright blocking or demonetization. YouTube is not the only platform to leverage demonetization as a form of moderation, however it is the largest and most well known and our study focuses on it because it provides a near-unique environment to study this phenomenon. As debates surrounding moderation on platforms have started to receive public and regulatory attention, it is important that lawmakers and regulators recognize demonetization as one of its forms. Particularly relevant to the demonetization and moderation processes of online platforms is §230 of the Communications Decency Act [4] in the United States which currently grants blanket immunity to platforms for publishing, promoting, and censoring users' speech. We expect any changes to this blanket immunity to have a profound impact on the underlying monetization, moderation, and recommendation algorithms leveraged by platforms. It currently remains unclear in which direction regulation in the United States will move – i.e., removing platform protections for not adequately censoring problematic speech (as is the case in the European Union [5]), or removing platform protections for censoring speech (as has been proposed by former US President Trump [31]). Regardless of the direction taken, it is becoming increasingly clear that the *effective enforcement* of any new policies, either by private right of action or regulatory bodies, will require added transparency in the way of *preservation of and access to interpretable* algorithmic decisions and their associated meta-data. This is critical because even extensive studies which measure opaque decisions from external vantage points are fundamentally limited in their ability to provide satisfying explanations for algorithmic decisions and will consequently fail to provide any actionable data for regulatory bodies or individual parties to act on.

Impact of algorithmic opacity on research. Extending beyond regulatory bodies, the absence of transparency in algorithmic decision making also impact researchers seeking to better understand the consequences of our reliance on AI. The need for researcher access to algorithmic decisions is particularly highlighted by the fact that limitations faced by our study could have been avoided by researcher access to algorithmic demonetization decisions by YouTube. More specifically, the unavailability of a clear monetization signal required us to improvise an expensive heuristic (requiring five page loads for each video to check for the presence of ads) to infer monetization status. Further complicating matters, YouTube's refusal to loosen restrictions on the rate at which our IP addresses could request videos negatively impacted the granularity at which our heuristic could observe changes in monetization status. Similar challenges have been observed in other studies seeking to audit algorithmic decisions (e.g., [54] which also calls for increased transparency in YouTube comment moderation decisions). This lack of transparency results in the need for improvised signals and high-frequency measurements whose rate is ultimately controlled by the platform. Evident through the platform-imposed limits on this research and other efforts to audit large online platforms [52] is the existence of a significant power asymmetry between online platforms that have a significant role in shaping real-world discourse and events and the researchers seeking to audit their algorithms to verify fairness, identify explanations for decision making, and to verify their adherence to public statements and policies. As the power and ubiquity of platforms and algorithmic decision-making continue to grow, it is critical that researchers are empowered with the ability to study them.

Impact of algorithmic opacity on creator-platform relationships. The controversies surrounding platform demonetization policies and decisions arise from the lack of transparency in current algorithmic decision making which consequently leads to widespread accusations of organizational biases against specific types of content or content creators [3, 85]. It is important to note that this lack of transparency extends not just to the public, but also to the creators that contribute to and generate revenue for platforms. For example, in the case of YouTube, creators are not made aware of when or why their content is demonetized. A consequence of this information asymmetry that exists between creators and the platforms they create for is increased creator frustration, resentment, anxiety, and the perception of algorithmic gatekeeping [92]. These negative feelings are only exacerbated when creators who are heavily reliant on advertising revenue from the platform attempt and fail to mold their content to avoid their perceived understanding of the triggers for algorithmic demonetization. [32, 61]. Increasingly, these challenges have led creators to resort to alternate revenue streams which rely on direct support from consumers (e.g., by selling merchandise and developing content exclusively for Patreon supporters) or in-content ads and product placements [32]. This diversification has the potential to negatively impact platforms. First, as our study highlights, the growth of smaller creators (under 1M subscribers) are already disproportionately impacted by these algorithmic decisions and the impact on their revenue is only increased due to their inability to generate alternate revenue streams — potentially reducing their future commitment to creating for the platform. Second, as more popular creators continue to reduce their reliance on advertising revenue, creating "advertiser-friendly" content may become a second-order priority — potentially impacting the amount of content on which platforms may generate revenue from.

9 CONCLUSIONS

Given the lack of transparency surrounding demonetization decisions made by YouTube, despite its limitations, our study sheds light on several important characteristics of the algorithm's behavior and effect.

Demonetization of a video is a relatively infrequent occurrence (§4). In our dataset, which consists of over 354K videos whose creators were sampled from controversial (banned subreddits) and mainstream (YouTube Trending) sources, only 0.5% (1,810) of all videos ever experience a demonetization transition. What is notable, however, is that over a quarter of these (484) are eventually remonetized — suggesting corrections to previous demonetization decisions made by the algorithm are frequent. Breaking results down by the source of ingestion, we see that the rates of non-monetization and de-monetization are in fact dependent on the source, with controversial creators having a higher likelihood of non-monetization and lower likelihood of remonetization when compared with mainstream creators.

Time to arrive at a final monetization decision can be lengthy (§4). Our analysis shows that the time taken to arrive at a final demonetization or remonetization decision can require several days, despite the use of AI for decision making. The median time to demonetization for videos in our dataset is found to be 116 hours and the median time to remonetization is found to be 129 hours. This difference of 13 hours suggests the time required for additional review following an appeal of the initial demonetization decision. When considering ingestion sources, we find that there is no statistical differences between the times to demonetization and remonetization based on whether creators were found on controversial or mainstream sources. These findings suggest that although the rates of de- and re-monetization are different across these groups, the process by which these monetization decisions are made are similar.

Larger channels benefit from more favorable demonetization and remonetization rates (§5.1). Our study shows the presence of a statistically significant lower demonetization and higher remonetization rate for large channels with over 1M subscribers when compared with smaller channels. This finding lends credence to claims of platform bias and tiered governance structures that favor larger creators [32]. Our analysis also shows that this favorable treatment extends to the remonetization process with larger channels experiencing a statistically significant and shorter time to remonetization than smaller channels. This finding is supported by YouTube's own statement that videos likely to gain substantial traffic are a priority for human reviewers [7].

Keyword occurrences cannot explain demonetization decisions (§5.2). In our analysis of demonetization rates conditioned on word-type occurrences, we found evidence to suggest that demonetization decisions are *not* based on the occurrence of keywords in video titles. This finding contradicts previous claims of the existence of a keyword blacklist [38] and suggests the use of a more nuanced algorithm.

Many video topics with high demonetization rates are explainable, however, concerns remain about the absence of bias (§5.2). Our analysis on the relationship between video content and demonetization rates found no evidence of political ideology-based biases in demonetization decisions. Instead, for videos in our dataset, demonetization decisions appeared to have been driven by the presence of copyrighted content such as sports broadcasts and popular films. We note, however, that videos related to the Islamic religion were found to have anomalously high rates of demonetization and low rates of remonetization when compared with the baseline — suggesting the possible existence of problematic biases that need to be investigated further.

Demonetization of a video appears to influence the growth and revenue of a channel (§6). For videos in our dataset, we find a statistically significant negative effect (-11.8 pp) on the video view rate for channels whose videos have been subject to demonetization. This effect is found to be most prominent for channels with between 100K and 1M subscribers (-30.1 pp). Given that creator revenue is impacted by number of views, these effect sizes are effectively a proxy for the revenue lost by demonetization. Our findings support previous work [32, 61] which suggest that the demonetization algorithm operates in tandem with the recommendation algorithm in order to maximize ad-revenue for the platform.

Challenges for future work. Our study focused on data gathered from a six-week period from July to September in 2020 – a period during which YouTube relied more heavily on AI-tools to enforce their monetization policies which remained unchanged. Given the frequent changes to monetization policies and the mechanisms to enforce them, there is an obvious need for long-term measurements to understand the nature and impact of changes to creators. There are challenges, beyond the obvious infrastructure costs, facing researchers who are considering these measurements, unfortunately. In November 2020, YouTube updated its monetization policies to allow them to display ads on any video – even content that was not monetized by the creator or from non-members of the YPP. In effect, this removes our ability to leverage the presence and absence of ads as a signal for monetization decisions. However, such measurements could still be used as a proxy to understand what content YouTube's algorithms deem "advertiser-friendly".

REFERENCES

[1] 2018. The forever war of PewDiePie, YouTube's biggest creator. https://www.washingtonpost.com/technology/2018/12/20/forever-war-pewdiepie-youtubes-biggest-creator/

[2] 2018. Top YouTube content categories by share of uploads 2018: Statista. https://www.statista.com/statistics/1026914/global-distribution-youtube-video-content-by-category/

[3] 2019. Divino Group LLC v. Google LLC. https://www.classaction.org/media/divino-group-et-al-v-google-llc-et-al.pdf

[4] 2020. 47 U.S. Code §230 - Protection for private blocking and screening of offensive material | U.S. Code | US Law | LII / Legal Information Institute. https://www.law.cornell.edu/uscode/text/47/230

[5] 2020. The Code of conduct on countering illegal hate speech online. https://ec.europa.eu/commission/presscorner/detail/en/qanda$_2$0$_1$135

[6] 2020. Wayback Machine: Advertiser-friendly content guidelines. https://web.archive.org/web/20200811083202/https://support.google.com/youtube/answer/6162278?hl=en&ref$_t$opic=9153642

[7] 2020. WayBack Machine: Request human review of videos marked "Not suitable for most advertisers" - YouTube Help. https://web.archive.org/web/20201020204246/https://support.google.com/youtube/answer/7083671?hl=en&ref$_t$opic=1115890

[8] 2020. Wayback Machine: Uploading videos to monetize with ads. https://web.archive.org/web/20200828082433/https://support.google.com/youtube/answer/75619389153642

[9] 2020. Wayback Machine: YouTube channel monetization policies - YouTube Help. https://web.archive.org/web/20200724232124/https://support.google.com/youtube/answer/1311392

[10] 2020. WayBack Machine: YouTube Community Guidelines & Policies - How YouTube Works. https://web.archive.org/web/20200811044808/https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#enforcing-policies

[11] 2020. Wayback Machine: YouTube Partner Program overview & eligibility - YouTube Help. https://web.archive.org/web/20200725004834/https://support.google.com/youtube/answer/72851?hl=en

[12] 2020. Wayback Machine: YouTube Rules and Policies: Community Guidelines. https://web.archive.org/web/20200725011550/https://www.youtube.com/howyoutubeworks/policies/community-guidelines/

[13] 2020. Wayback Machine: YouTube Rules and Policies: Monetization Policies. https://web.archive.org/web/20200722015509/https://www.youtube.com/howyoutubeworks/policies/monetization-policies/

[14] 2021. BRIEF OF AMICI CURIAE KYRATSO KARAHALIOS, ALAN MISLOVE, CHRISTIAN W. SANDVIG, CHRISTOPHER WILSON, FIRST LOOK MEDIA WORKS, THE AMERICAN CIVIL LIBERTIES UNION, THE AMERICAN CIVIL LIBERTIES UNION OF THE DISTRICT OF COLUMBIA, UPTURN, AND THE KNIGHT FIRST AMENDMENT INSTITUTE IN SUPPORT OF PETITIONER. https://cbw.sh/static/pdf/amicus$_a$clu.pdf

[15] 2021. Upcoming and recent ad guideline updates - YouTube Help. https://support.google.com/youtube/answer/9725604?hl=en&ref$_t$opic=9153642

[16] 2021. Van Buren is a Victory Against Overbroad Interpretations of the CFAA, and Protects Security Researchers. https://www.eff.org/deeplinks/2021/06/van-buren-victory-against-overbroad-interpretations-cfaa-protects-security

[17] 2021. Van Buren vs. United States. https://www.supremecourt.gov/opinions/20pdf/19-783$_k$53l.pdf

[18] 2021. YouTube channel monetization policies: How we enforce YouTube monetization policies - YouTube Help. https://support.google.com/youtube/answer/1311392?visit$_i$d=637535273191990793-852634117&rd=1#zippy=%2Cfollow-the-youtube-community-guidelines%2Cturn-off-ads-from-your-content%2Csuspend-your-participation-in-the-youtube-partner-program%2Csuspend-or-even-terminate-your-youtube-channel%2Chow-well-inform-you-of-actions-that-affect-your-monetization

[19] 2021. YouTube Data API: Google Developers. https://developers.google.com/youtube/v3

[20] 2021. YouTube Partner Program overview & eligibility - YouTube Help. https://support.google.com/youtube/answer/72851?hl=en#zippy=%2Cim-no-longer-in-ypp-or-i-was-never-in-the-program-and-im-seeing-ads-on-my-videos-am-i-earning-revenue-from-those-ads%2Cwhat-if-i-dont-meet-the-program-threshold

[21] Alberto Abadie. 2005. Semiparametric difference-in-differences estimators. *The Review of Economic Studies* 72, 1 (2005), 1–19.

[22] Ali Alkhatib. 2021. To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA.

[23] Ali Alkhatib and Michael Bernstein. 2019. Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[24] Mike Ananny. 2018. *Networked press freedom: Creating infrastructures for a public right to hear.* MIT Press.

[25] Mark Andrejevic. 2009. Exploiting YouTube: Contradictions of user-generated labor. *The YouTube Reader* 413 (2009), 36.

[26] Joshua D Angrist and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

[27] Camila Souza Araújo, Gabriel Magno, Wagner Meira Jr., Virgílio A. F. Almeida, Pedro Hartung, and Danilo Doneda. 2017. Characterizing Videos, Audience and Advertising in Youtube Channels for Kids. In *Social Informatics - 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I (Lecture Notes in Computer Science)*, Giovanni Luca Ciampaglia, Afra J. Mashhadi, and Taha Yasseri (Eds.), Vol. 10539. Springer, 341–359. https://doi.org/10.1007/978-3-319-67217-5$_2$1

[28] David Bamman, Brendan O'Connor, and Noah A. Smith. 2012. Censorship and Deletion Practices in Chinese Social Media. *First Monday* 17, 3 (2012).

[29] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 830–839.

[30] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.

[31] Abram Brown. 2020. What Is Section 230—And Why Does Trump Want To Change It? https://www.forbes.com/sites/abrambrown/2020/05/28/what-is-section-230-and-why-does-trump-want-to-change-it/?sh=552684b2389d

[32] Robyn Caplan and Tarleton Gillespie. 2020. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. *Social Media + Society* 6, 2 (2020), 2056305120936636. https://doi.org/10.1177/2056305120936636 arXiv:https://doi.org/10.1177/2056305120936636

[33] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[34] Le Chen, Alan Mislove, and Christo Wilson. 2015. Peeking Beneath the Hood of Uber. In *Proceedings of the Internet Measurement Conference (IMC 2015)*. Tokyo, Japan.

[35] Le Chen, Alan Mislove, and Christo Wilson. 2016. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *Proceedings of the 25th International World Wide Web Conference (WWW 2016)*. Montreal, Canada.

[36] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1217–1230.

[37] Thomas J DiCiccio, Bradley Efron, et al. 1996. Bootstrap confidence intervals. *Statistical science* 11, 3 (1996), 189–228.

[38] Lindsay Dodgson. 2019. YouTubers have identified a long list of words that immediately get videos demonetized, and they include 'gay' and 'lesbian' but not 'straight' or 'heterosexual'. https://www.insider.com/youtubers-identify-title-words-that-get-videos-demonetized-experiment-2019-10

[39] Elizabeth Dwoskin. 2019. YouTube's arbitrary standards: Stars keep making money even after breaking the rules. *The Washington Post* (2019).

[40] Yelena Dzhanova. 2019. Forget law school, these kids want to be a YouTube star. https://www.cnbc.com/2019/08/02/forget-law-school-these-kids-want-to-be-a-youtube-star.html

[41] Bradley Efron. 1987. Better bootstrap confidence intervals. *Journal of the American statistical Association* 82, 397 (1987), 171–185.

[42] Jacob Eisenstein. 2013. What to do about bad language on the internet. In *Proceedings of the 2013 conference of the North American Chapter of the association for computational linguistics: Human language technologies*. 359–369.

[43] Jacob Eisenstein. 2019. *Introduction to Natural Language Processing*. MIT Press.

[44] Megan Graham. 2020. YouTube will put ads on non-partner videos but won't pay the creators. https://www.cnbc.com/2020/11/19/youtube-will-put-ads-on-non-partner-videos-but-wont-pay-the-creators.html

[45] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[46] Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, suppl 1 (2004), 5228–5235.

[47] Justin Grimmer and Brandon M Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis* 21, 3 (2013), 267–297.

[48] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In *Proceedings of the 22nd International World Wide Web Conference (WWW 2013)*. Rio de Janeiro, Brazil.

[49] Miguel A Hernán and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

[50] Margaret Holland. 2016. How YouTube developed into a successful platform for user-generated content. *Elon Journal of Undergraduate Research in Communications* 7, 1 (2016).

[51] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.

[52] Knight Institute. 2021. Researchers, NYU, Knight Institute Condemn Facebook's Effort to Squelch Independent Research about Misinformation. https://knightcolumbia.org/content/researchers-nyu-knight-institute-condemn-facebooks-effort-to-squelch-independent-research-about-misinformation

[53] Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Bias Misperceived:The Role of Partisanship and Misinformation in YouTube Comment Moderation. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 01 (Jul. 2019), 278–289. https://ojs.aaai.org/index.php/ICWSM/article/view/3229

[54] Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2020. Reasoning about Political Bias in Content Moderation. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 09 (Apr. 2020), 13669–13672. https://doi.org/10.1609/aaai.v34i09.7117

[55] Shan Jiang and Christo Wilson. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[56] Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc.

[57] Jacob Kastrenakes. 2020. YouTube will rely more on AI moderation while human reviewers can't come to the office. https://www.theverge.com/2020/3/16/21182011/youtube-ai-moderation-coronavirus-video-removal-increase-warning

[58] Katherine Keith, David Jensen, and Brendan O'Connor. 2020. Text and Causal Inference: A Review of Using Text to Remove Confounding from Causal Estimates. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 5332–5344.

[59] Martin Kenney and John Zysman. 2016. The Rise of the Platform Economy. *Issues in science and technology* 32 (03 2016), 61–69.

[60] Emre Kiciman, Scott Counts, and Melissa Gasser. 2018. Using longitudinal social media analysis to understand the effects of early college alcohol use. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[61] Sangeet Kumar. 2019. The algorithmic dance: YouTube's Adpocalypse and the gatekeeping of cultural content on digital platforms. *Internet Policy Review* 8, 2 (2019), 1–21. https://doi.org/10.14763/2019.2.1417

[62] Liz Lavaveshkul. 2012. How to achieve 15 minutes (or more) of fame through youtube. *J. Int'l Com. L. & Tech.* 7 (2012), 370.

[63] Paige Leskin. 2019. American Kids Want to Be YouTube Stars: Survey. https://www.businessinsider.com/american-kids-youtube-star-astronauts-survey-2019-7

[64] Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028* (2002).

[65] Rebecca MacKinnon. 2009. China's censorship 2.0: How companies censor bloggers. *First Monday* 14, 2 (February 2009).

[66] Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. (2002). http://mallet.cs.umass.edu.

[67] McKeown. 2020. Prager University v. Google LLC: Opinion from Judge McKeown. https://cdn.ca9.uscourts.gov/datastore/opinions/2020/02/26/18-15712.pdf

[68] Toby Miller. 2009. *Cybertarians of the world unite: You have nothing to lose but your tubes!* National Library of Sweden.

[69] Stephen L Morgan and Christopher Winship. 2015. *Counterfactuals and causal inference*. Cambridge University Press.

[70] Dong Nguyen, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and Jane Winters. 2020. How we do things with words: Analyzing text as social and cultural data. *Frontiers in Artificial Intelligence* 3 (2020), 62.

[71] David B Nieborg and Thomas Poell. 2018. The platformization of cultural production: Theorizing the contingent cultural commodity. *New media & society* 20, 11 (2018), 4275–4292.

[72] Brendan O'Connor, David Bamman, and Noah A Smith. 2011. Computational text analysis for social science: Model assumptions and complexity. In *Second workshop on comptuational social science and the wisdom of crowds (NIPS 2011)*. Citeseer.

[73] Kostantinos Papadamou, Antonis Papasavva, Savvas Zannettou, Jeremy Blackburn, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Michael Sirivianos. 2020. Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. *Proceedings of the International AAAI Conference on Web and Social Media* (2020).

[74] Kostantinos Papadamou, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. 2021. Understanding the Incel Community on YouTube. arXiv:cs.CY/2001.08293

[75] Judea Pearl. 2009. *Causality*. Cambridge university press.

[76] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[77] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50.

[78] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira Jr. 2020. Auditing radicalization pathways on YouTube. In *FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020*, Mireille Hildebrandt, Carlos Castillo, Elisa Celis, Salvatore Ruggieri, Linnet Taylor, and Gabriela Zanfir-Fortuna (Eds.). ACM, 131–141. https://doi.org/10.1145/3351095.3372879

[79] Caitlin M Rivers and Bryan L Lewis. 2014. Ethical research standards in a world of big data. *F1000Research* 3 (2014).

[80] Margaret E Roberts. 2018. Censored: Distraction and Diversion Inside China's Great Firewall. (2018).

[81] Margaret E Roberts, Brandon M Stewart, and Richard A Nielsen. 2020. Adjusting for confounding with text matching. *American Journal of Political Science* 64, 4 (2020), 887–903.

[82] Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science* 58, 4 (2014), 1064–1082.

[83] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* (2014).

[84] Laura Sehl and Philippe Ross. 2020. Demonetization on YouTube and the Visibility of News Produced by Non-Mainstream News Commentators.

[85] Erin Shaak. 2020. Racism for Profit: Black Content Creators Sue YouTube Over Discriminatory Access Restrictions. https://www.classaction.org/blog/racism-for-profit-black-content-creators-sue-youtube-over-discriminatory-access-restrictions

[86] SocialBlade. 2021. SocialBlade: Analytics Made Easy. https://socialblade.com

[87] Stephen Stanford. 2018. YouTube and the Adpocalypse: How Have The New YouTube Advertising Friendly Guidelines Shaped Creator Participation and Audience Engagement?

[88] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[89] Geoff Weiss. 2017. Creators Cry Foul After YouTube Demonetizes Casey Neistat's LoveArmyLasVegas Video - Tubefilter. https://www.tubefilter.com/2017/10/06/youtube-demonetization-casey-neistats-love-army-las-vegas/

[90] Wayne W Wilkinson and Stephen D Berry. 2020. Together they are Troy and Chase: Who supports demonetization of gay content on YouTube? *Psychology of Popular Media* 9, 2 (2020), 224.

[91] Rolf Winkler, Jack Nicas, and Ben Fritz. 2016. PewDiePie Says WSJ Took Anti-Semitic Content Out of Context. https://www.wsj.com/articles/pewdiepie-says-wsj-took-anti-semitic-content-out-of-context-1487278375

[92] Eva Yiwei Wu, Emily Pedersen, and Niloufar Salehi. 2019. Agent, Gatekeeper, Drug Dealer: How Content Creators Craft Algorithmic Personas. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 219 (Nov. 2019), 27 pages. https://doi.org/10.1145/3359321

[93] Leon Yin and Aaron Sankin. 2021. Google Has a Secret Blocklist that Hides YouTube Hate Videos from Advertisers—But It's Full of Holes – The Markup. https://themarkup.org/google-the-giant/2021/04/08/google-youtube-hate-videos-ad-keywords-blocklist-failures

[94] YouTube. 2020. Protecting our extended workforce and the community. https://blog.youtube/news-and-events/protecting-our-extended-workforce-and/

[95] YouTube. 2021. Trending on YouTube. https://support.google.com/youtube/answer/7239739?hl=en

[96] YouTube. 2021. YouTube for Press: YouTube in numbers. https://www.youtube.com/intl/en-GB/about/press/

[97] YouTube. 2021. YouTube Partner Program overview & eligibility - YouTube Help. https://support.google.com/youtube/answer/72851?hl=en

[98] Savvas Zannettou, Sotirios Chatzis, Kostantinos Papadamou, and Michael Sirivianos. 2018. The Good, the Bad and the Bait: Detecting and Characterizing Clickbait on YouTube. In *2018 IEEE Security and Privacy Workshops (SPW)*. 63–69. https://doi.org/10.1109/SPW.2018.00018