YouTube Platform Interventions discussed in Videos about Climate Change Skepticism

Seminar "Critical social media analysis using mixed methods", Winter Term, 2020/21

Isabel Schmuck
Technische Universität Berlin
Berlin, Germany
i.schmuck@campus.tu-berlin.de

Rahaf Gharz Addien
Freie Universität Berlin
Berlin, Germany
rahag65@zedat.fu-berlin.de

Julius Tembrockhaus
Freie Universität Berlin
Berlin, Germany
julius.tembrockhaus@fu-berlin.de

ABSTRACT

The aim of this study is to investigate how different types of platform interventions are discussed in the comment sections of YouTube videos in the context of climate change skepticism. The work focuses on exposing explanation and justification patterns to real and perceived platform interventions, determining the volume of comments debating the intervention methods as well as the user's opinion on the relationship between interventions and free speech, with particular focus on platform credibility. An elaboration of the underlying theoretical background combined with a detailed literature review was carried out to accurately highlight the necessities, implementation strategies and consequences of interventions on social media platforms.

To answer the research question a preliminary quantitative analysis followed by an extensive qualitative analysis was performed. The comment sections of ten precisely selected YouTube videos were considered where the average proportion of comments actually dealing with platform interventions was determined to be 3% corresponding to an average number of 99 comments of interest per video.

A central component of the criticism on platform interventions consists in complaints about the YouTube search algorithm and the accessibility of specific videos via the search function. This resentment becomes especially evident by the observable trend of moving controversial content and entire communities to other platforms. It is further noticeable that the commenting users have fundamentally different perceptions of YouTube's role as a social platform. A considerable part of the users understand intervention methods as a form of censorship and an infringement against their right of self-expression and accuse YouTube to follow a hidden agenda. In contrast, other users accept and justify the execution of platform interventions and request to delete unacceptable content more quickly. Overall, a positive correlation between the evaluation of the platform's credibility and the evaluation of platform interventions was observable.

1 INTRODUCTION

Since the first video was published on YouTube in 2007, the concept of how people navigate the platform including what kind of videos they uploaded has changed drastically. Short funny clips no longer dominate the platform and have not done so for a while. The platform's rise in popularity comes hand in hand with a diversification of content and audiences. As a response, YouTube's moderation policies, as well as its search algorithm have been altered and exchanged many times. This development is partly due to technological advances, the changing relevance of social media to our daily lives, and as a direct response to the criticism. With professionally produced content that could easily be broadcasted on television and a rise of self-made YouTube personalities, YouTube has become a home for people wanting to explore alternative answers to contested topics, such as climate change.

In the light of this, YouTube came under criticism for directing its users to conspiracies and videos presenting false information on the causes and effects of climate change. For a long time, users looking at ordinary newscasts on a tropical storm would be channeled through recommended videos to content with ever increasing extremity ending up at a video climate change has been a government-driven hoax all along. Generally speaking, YouTube's search results depend on relevance calculated through criteria, such as the number of views, the number of likes, the right keywords and tags, and general user engagement (likes, comments and watch time). However, under the increasing pressure and with the changing political climate, YouTube introduced a number of counter measures to stop the spread of conspiracies, hateful and inflammatory content. Not only are some videos not shown as search results anymore but the appearance of informative banners and correctives was widely noted. Of course, these platform interventions are not uniformly welcomed. The comment sections of some videos are brimming with users discussing bans, banners, and the disappearance of certain videos. This led us to wonder if we can look at the users broaching platform interventions as a community with a common motivation, interest, or thought. Our focus, hereby, is not on the moderation of comments or specific moderation guidelines but user engagement with perceived bias against climate change dissent and the platform interventions by YouTube. We argue that the users we encounter in relevant comment sections relate platform interventions back to questions of censorship and free speech and, further, question the credibility of the platform when confronted with individual moderation attempts.

This seminar project aims to answer the question of how different types of platform interventions are discussed in videos about climate change skepticism on YouTube. Videos about climate change evoke a wide controversy between the users, particularly the videos of climate change skeptics. This makes climate change an appropriate topic for studying the discussion on platform interventions. We selected a number of videos on the topic of climate change, mostly of skepticism towards and rejection of climate change. We performed a quantitative analysis on the videos to calculate the proportion of comments debating interventions and find relevant clusters of comments for analysis. A qualitative approach is applied to derive the main arguments of users advocating for or arguing against interventions by the platform. The research questions we propose to guide our investigation and an outline of how we aim to answer them, are to be found in the following sections.

1.1 Research Questions

The main research question and the one that has guided our research project is: How are platform interventions discussed in videos of climate change skepticism? To help us in answering the main research question, we have identifies a couple of sub-questions. They progressively zoom into the arguments users make when discussing platform interventions and content moderation, in general. The first one is concerned with the volume of comments related to platform interventions. We ask, how many of the comments are dealing with platform interventions? The second one is concerned with a broader explanation and justification patterns to real and perceived platform interventions. Lastly, we ask how users are relating platform interventions to free speech and the credibility of the platform?

1.2 Outline

The paper is structured as follows: We start by giving an overview over the theoretical background. Thereby, we focus especially on virtual communities and why they would need moderation. Subsequently, we examine relevant research into moderation and its reception in regards to credibility and neutrality of the social media platform in the literature review section. In the methodology section, we introduce our approach to the selection of videos, quantitative and qualitative analyses. The results of our video selection, a short description of exemplary videos, a numeric breakdown of the proportion of comments dedicated to platform interventions for each video and individual arguments and topics we found in the comment sections are described in the results. Finally, we discuss our findings, outline the limitations of the project, and give an outlook.

2 THEORETICAL BACKGROUND

The following section provides an overview over the theoretical groundwork necessary for the analysis. The section will cover virtual communities and why they are moderated, which approach a community can take to moderation, content moderation beyond the removal of comments, platform interventions, and some theoretical background to censorship and social media platforms.

2.1 Virtual Communities and Moderation

The YouTube comment section is a space where users can share thoughts and debate issues. Users either deliberately chose to engage with the topic of the video or were funneled to the video through the YouTube video recommendation and search algorithms. When the topic of the video is controversial, such as commentary on politics or religion, these discussions can be hateful and inflammatory. However, even local Facebook groups centered around fermentation best practices or podcast recommendations can go up in metaphorical flames if left unattended for too long. Moderators play an important role in any online forum by helping to maintain order and facilitating discussion within the community. Generally, a community can be comprised of a handful of people but also as big as the internet itself. Communities can overlap and they can nest. Useful for talking about virtual communities is the definitions of three of their key elements: "the community's members, the content they share with each other, and the infrastructure they use to share it" (p. 48) [6]. YouTube's comment section can be understood as the infrastructure, its' users as the members, and the content of the comments in the form of text, links, tags, etc. are the content. Generally, moderation has three big goals: producing a productive environment where information is generated and exchanges, facilitating openness and access to the community and its resources, and, finally, to do its job at a low cost, which means, making demands on the community's members and infrastructure as few as possible. Most moderation of virtual communities exists on a spectrum of these goals and trade-offs between them are evident [6].

Summarizing the objective of moderation, Grimmelmann (2015) writes: "The interface between infrastructure and information is vulnerable to some predictable forms of strategic behavior, including spam, harassment, and other famous pathologies of online life. These are the abuses against which moderation must guard. Moderation need not prevent them entirely—and probably cannot without killing the commons—but it must keep them within acceptable bounds, and without driving up the costs of moderation itself to unacceptable levels"(p. 53) [6]. In the last couple of years, some of those "famous pathologies of online life" have become increasingly relevant in making social media platforms come under heightened scrutiny. Platforms are criticized not only for their well-documented privacy issues, but also for how they are used to spread misinformation, breed violent hate speech, and promote partisanship [8].

2.2 Approaches to Moderation

To address dangers to their virtual communities and combat hate and misinformation, social media platforms engage in moderation to enforce community guidelines and other boundaries for how users behave and what they can say. When we talk about moderation we borrow Grimmelmann's (2015) definition of moderation as "the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse" (p. 6) [6]. Implicit in this definition is a quasi-governing entity that decides both on these mechanisms as well as the kind of participation to facilitate and the kind of abuse to prevent. The governing body in the case of a video comment section is the social media platform attached. However, content moderation for social media platforms is not a one-size-fits-all affair. Most social media platforms underlie similar commercial and legal restrictions. Still, there are vastly different ways of approaching moderation. Myers West (2018) argues that there are two ways of approaching moderation depending on the framing of the platform. There is the "free speech" oriented company and the "community" oriented company. The first would probably organize community guidelines around defending the expression of free speech, while the second is more likely to place importance on fostering good behavior and culling harmful behavior among its users. The biggest examples of these two orientations would be Twitter and Facebook, respectively [16].

Additionally relevant to the differences in approaching content moderation is the makeup of a social media company. A company like YouTube is not a homogenous monolith but a fragmented, heterogenous organization with many actors working in different directions simultaneously. These organizations are made up of many “teams of employees that may have different cultures, values, and incentives with regard to public expression that collectively may or may not have coherence” (p. 4367) [16]. In that sense, the team within YouTube concerned with garnering advertising revenue might have very different ideas on the issue as the team organized around creating community guidelines.

2.3 Content Moderation beyond the Removal of Comments

There are many ways to approaching moderation practically. One mechanism is to hire human moderators to go over platform-specific content. Another mechanism is to invest in Bayesian filters to recognize skin and spam or partner with fact-checking organization to identify and contest misinformation. On a policy level, social media platforms have updated and expanded their terms of service and community guidelines outlining what constitutes hate speech, harassment, or misinformation. In the case of YouTube, the community guidelines “list[s] rules for: nudity or sexual content, harmful or dangerous content, hateful content, violent or graphic content, harassment and cyberbullying, etc.” (p. 279) [8]. If a video or comment is flagged as violating the community guidelines it is moderated, i.e. moderated. In some cases, moderation of content can have repercussions for the content’s author, such as being banned from the platform for a set amount of time.

The kind of content we are interested in specifically are comments under YouTube videos. In the case of the moderation of comments, several reasons can cause a comment to be removed from YouTube. The comment might have been reviewed by human YouTube moderators scanning many comments. The comment might have been individually flagged by a YouTube user and then reviewed by official YouTube moderation. Finally, the comment in question could have been removed either by the author of the comment themselves or by the author of the video the comment was under. In addition to human efforts, YouTube also employs algorithmic solutions to the issue. These algorithms automatically flag and moderate inappropriate content. In general, the mechanisms that lead to comment moderation are not clear cut and employed only in specific instances but convoluted [8].

2.4 Platform Interventions

However, the moderation and removal of comments and videos is not the only measure that social media platforms can take to combat hate speech and foster a productive community. In addition to moderating comments, the platform can also engage in what we call platform interventions. A platform that allows its users to make money with their content can attempt to de-monetize the particular user or channel. This would de-incentivize users to produce inflammatory content or behave otherwise in a way that would impact their ability to gain ad revenue off of the site. Platforms can also ban individual users. In addition to individual users, they can also suspend accounts and channels. In some cases, the ban is temporary and users are eventually allowed back on the site. In other cases, users are never allowed back on the site and would have to officially appeal to the platform to be let back on. Depending on the sophistication of the platform’s authentication mechanisms, these bans cannot be circumvented by just creating a new e-mail address [16]. A relatively new and contested strategy is shadow banning, also called stealth banning, which describes the practice of limiting the visibility of a user’s content compared to their regular visibility without their explicit knowledge thereof. To this day, it is still contested, if this is a practice that is actually happening and if yes, what kind of creators are impacted by it the most [14]. Finally, some platforms have begun adding informative banners to content most affected by deliberate misinformation campaigns and conspiracy theories. These banners lead to resources about the issue at hand, mostly government websites or Wikipedia. All of these mechanisms are additional measures to foster open and productive virtual communities.

2.5 Content Moderation and Free Speech

Of course, content moderation and platform interventions of any kind are not received without controversy and conflict. Moderation practices are often accompanied by cries of censorship and calculated suppression of a particular brand of opinions. In legal terms, illegitimate censorship is suppressing particular individuals, opinions, or forms of communication unfairly. The line between legitimate form of moderation to discourage harmful or inappropriate behavior and illegitimate censorship can appear hard to define and accept. „Censorship is usually defined subjectively, and in cases where there is room for interpretation, the unconscious biases of regulators may affect their judgments. On the other hand, a user’s own bias may lead them to perceive unfair treatment where there is none “(p. 350) [21]. As commercial platforms, social media organizations are private spaces. Their absorption of civic roles and de facto function as spaces for self-expression and “collective association” has little effect on their architecture and governance. They cannot be held to the same standards as legally “public” spaces. Social media platforms are “not simply about facilitating user-produced content across networks to large audiences or “endusers”; rather, they are primarily concerned with establishing the technocultural conditions within which users can produce content and within which content and users can be re-channeled through techno-commercial networks and channels“ (p. 4367) [16]. However, even just the suspicion of illegitimate censorship can create atmospheres where users feel mistreated and trust in a platform is undermined, regardless of whether claims of censorship are actually true. Fair moderation practices without illegitimate censorship is a basis for creating engaging online spaces for deliberation on controversial topics that are safe to all users. To keep the community open and productive, difficult decisions and careful deliberations about the scope and target of specific content moderation practices and platform interventions are necessary [21].

3 LITERATURE REVIEW

The following section introduces recent and ongoing research on YouTube algorithms and comment sections, as well as research on some of the platform interventions individually. Additionally, credibility as a concept, is investigated more closely.

In a 2018 article, Sarah Myers West interviewed users on how they have experienced content moderation [16]. Myers West focuses on how users of several platforms perceive being moderated and she contextualizes this by highlighting what information users have available to understand content guidelines. In her interviews, users express theories about how and why the content was flagged. Myers West reports that users develop “folk theories” about how platforms work generated out of each user’s personal experiences. The most common theories were other users’ having flagged their content. Additionally, users perceived the “flagging” as done by a disembodied “they” and as emblematic of a political bias in the platform. Users expressed the belief that social media companies should be held to the same standards regarding free speech as the US government, regardless of the reason for the content takedown and regardless of current regulations regarding social media platforms. Myers West reports: “Often, content moderation is described as a form of censorship or restraint on a user’s voice” (p. 4373). Users in the study overwhelmingly ascribed the moderation of their content to flagging by another user without considering technological or algorithmic interventions. Without much information about why the particular user was reported or how to appeal, Myers West concludes “the design of content moderation systems works at cross purposes with this intention: they make people feel confused, frustrated, and as though they are “shouting into a void”” (p. 4380) [16].

Jiang, et al (2019) examine the entanglement between content moderation and the perceived explicit biases in moderators quantitatively [8]. To do so, the researchers use YouTube to investigate how to string political affiliation and misinformation impact the likelihood of comment moderation. Jiang, et al use YouTube comments which were previously labeled according to partisanship and presence of misinformation. Since the researchers have no access to YouTube’s internal system, they focus on the outcome of moderation and how it is perceived. Running a basic correlational analysis, their findings suggest “a political bias against right-leaning content” (p. 279) [8]. However, they conclude the bias to be “misperceived” as the analysis ignores other factors making a ban likely, such as bullying. Once the confounding variables, such as linguistic features of the comments and social engagement, are controlled, they find no evidence of a political bias against right-leaning comments in the moderation of YouTube comments. Instead, “the greater amount of comment moderation on right-leaning videos is explained by correspondingly higher levels of misinformation, extreme partisanship of videos, and various linguistic signals (e.g., hate speech) in comments” (p. 287). In these cases, the moderation of a comment is entirely consistent with YouTube’s community guidelines [8].

In their analysis of YouTube as an influential source of information, Rieder, Matomoros-Fernández, and Coromina (2018) focus on search results of the YouTube search bar [19]. They argue that their presentation as an ordered list suggests an attribution of relevance and the idea that some content deserves more attention and prominence. “Through the visibility given to metadata such as view, like and comment counts, this competitive process is extended throughout the platform and becomes an influential element in information diffusion as users take such indicators into account when viewing and sharing” (p. 52) [19]. Considering both the technical implementation and the user engagement as input, the researchers call for understanding ranking as a complex process which unfolds over time. They use YouTube Data Tools [18] to make the same call on 44 consecutive days and analyzed the resulting list of videos concentrating on the top 20 videos sorted by relevance. Their findings show that the search function is reactive to attention cycles and dominated by YouTube-originated content. This ranking allows what the researchers call “niche entrepreneurs” to be prominent by feeding off of conflict and loyal audiences. Ultimately, they dispute the idea that the YouTube search results are purely a popularity metric [19].

There is also Lamerichs’ (2020) account of user tactics on YouTube and Tumblr [9]. Examples of Instagram, YouTube and Tumblr show that platforms are taking control over data flow on their platform. However, Lamerichs argues that strategies, such as banning certain hashtags, can have detrimental effects on certain user communities and user cultures, most often subcultures and minority groups. “Users are “de-platformed” in a sense that they suddenly cannot search for their community and peers” (p. 36). These decisions can result in enormous backlash. Lamerichs analyzes the way users respond to changes in policies and algorithms on Tumblr in 2018 and YouTube in 2019. While her focus is on the YouTube algorithm and how content creators try to game it, she finds that users are not fully susceptible to the whims of platform politics but deploy tactics to actively shape it. As a response to YouTube updating its recommendation system, YouTubers published video tutorials on how to “game the system” and reach as many people as possible with their content. However, many user tactics go beyond trusting in quantitative metrics and SEO. In that sense, YouTube users and creators are not just passively using a platform but debate and strategize ways to game the platform and its rules all the time [9]. There are pervasive changes to how humans consume information through social media versus traditional new outlets. This requires new skills in judging the credibility of a piece of information off its consumer. Information credibility is defined as the extent to which one perceives information to be believable [13]. In their research to determine factors that influence the information credibility on social media platforms, Ruohan Li and Ayoung Suh [11] establish three dimensions of information credibility which are medium, message, and source credibility. Medium credibility refers to the perceived levels of credibility a specific medium that individuals used, message credibility relies on the strength of the used arguments and the quality of the information, and source credibility focuses on the trustworthiness of the source as the likelihood to provide credible information.

The credibility of social media platforms as a medium is identified by three factors: the dependency of individuals (users) on the platform as a source of information and having no alternatives, the interactivity between the platform and the users where people tend to believe the content of videos that have high numbers of likes and views, and transparency which is related to an individual’ perceived willingness to share information freely and frankly with others on a specific medium [11].

4 METHODOLOGY

This section names, describes, and explains the methods used to evaluate the data of the study. The individual steps of the analysis are specified with regard to the theoretical background of the methodology.

4.1 Candidate Video Selection

To analyze debates on YouTube platform interventions, a lot of comments are needed. The key of this task is to select videos, where specific intervention strategies are actually discussed by the users within the comment sections. To make the data gathering reproducible and avoid personal bias in our video selection, we decided to implement an automatic pipeline. Another reason for the automation of the data gathering process is the difficulty to manually find climate change denying videos. By screening a larger number of videos, the chances to detect appropriate videos for our study increases.

The pipeline starts by using the *Video List Module* of YouTube Data Tools [18] to generate a data set of YouTube videos resulted by different search queries. We always requested the top 50 results for each query sorted by relevance and restricted the time frame. The three search query strings as well as the corresponding parameters are displayed in Table 1. We further defined a list of keywords that are indicators for the presence of a discussion about platform interventions in the respective comment sections. The entire list contains 23 elements and is documented in Table 2. Based on the three search queries a data set of 150 instances is created. For each video in the set the 200 most relevant comments are fetched via the YouTube Data API v3 implemented in python. If a comment has replies, they are also considered. Afterwards, a word count calculation is performed on the fetched comments for each video and if at least two of the defined keywords with a minimum of three occurrences are present in the data, a video is classified to be a candidate video for our study. Since in YouTube comments spelling mistakes are to be expected, we set the required word similarity to 0.9 and used to *get_close_matches()* function of the *difflib* library to prevent this problem. By following these thresholds 21 of 150 videos could be identified as candidates. For many of them the minimum numbers for present keywords are far exceeded. The *Video Info and Comments Module* of YouTube Data Tools [18] is then used to download the entire comment section for each candidate video. In the following, the data set of candidate videos is further analyzed to extract the comments debating on platform interventions from the remaining off-topic data.

4.2 Quantitative Analysis

During a preprocessing step the data is cleaned by dropping empty rows and unused columns to reduce the size of the data set. Furthermore, HTML-tags, links and usernames are removed to create anonymity for the users and focus on the content of the comments. Since our study aims to highlight the main arguments and characterize the general opinions in the discussion about platform intervention, we are not interested to trace the arguments back to individual users. Afterwards, text embeddings of the comments are generated by the Universal Sentence Encoder v5 [2] of Google's TensorFlow Hub, which was pre-trained with a Transformer encoder structure. Here, the comments are encoded into high dimensional vectors that are used to perform subsequent tasks. The encoder model was optimized for greater-than-word length texts, such as sentences, phrases or short paragraphs [2] and therefore it is perfectly applicable to encode YouTube comments. The output of each encoded comment is then a 512 dimensional vector.

To compare the comments in terms of semantics, a similarity analysis of the vector-based embeddings is performed by calculating the cosine of the angles between the embedding vectors [22]. The cosine similarity is a measure between non-zero vectors that is widely used to detect text similarity. The cosine of the angle between two vectors is the same as the inner product of the same vectors and return a score between 0 and 1. Hereby 1 refers to as exact similar and 0 specifies no similarity.

In the next step, the comments are clustered by the KMedoids algorithm based on their similarity with the aim to mainly focus on groups of similar comments related to the topics of platform interventions. In the algorithm a medoid is defined as the point in each cluster, whose dissimilarity with all other data points in the cluster is minimal. Since we do not expect non-spherical data groups in our similarity analysis, KMedoids' concept of minimizing distances between non-medoid objects and the medoids can be applied without any problems. The algorithm is simple, converges in a fixed number of steps and is less sensitive to outliers (most likely present in the similarity calculations of YouTube comments) compared to other partitioning algorithms. Therefore KMedoids provides a good fit for our clustering method.

However, prior to the clustering the dimensions of the text embeddings need to be reduced and the optimal number of clusters has to be determined. The originally 512 dimensions of the comment embedding vectors are decreased to 100 dimension by the UMAP algorithm [12] to make possibly present clusters more pronounced. Due to the fact that KMedoids takes the number of clusters as an input to generate the respective number of initial medoids, the best k number of clusters need to be known before the actual clustering. Therefore, we precalculate the clustering with $5 < k < \frac{\text{\# comments}}{10}$ in steps of 5 and assess the silhouette scores of the clustering results. The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). It ranges from -1 to 1 and high values indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. Based on these calculations the best k number of clusters is defined and the actual clusering is performed.

When the clustering is completed, the most frequent words per cluster are calculated after a previous exclusion of stopwords. Here, we used the stopwords corpus provided by NLTK [17]. Iterating over the clusters, a candidate cluster is identified if at least one of the keywords (Table 2) can be found among the top 20 most frequent words in the cluster.

4.3 Qualitative Analysis

By investigating the candidate clusters the transition from the quantitative to a qualitative analysis takes place since the cluster examination needs to be performed manually. All comments in the candidate clusters that actually refer to platform interventions are

Table 1: YouTube Data Tools Input Queries

Search query	Iteration (items)	Timeframe	Rank by
"climate chnage hoax"	1 (50)	1970-01-01 to 2021-01-31	relevance
"climate chnage swindle"	1 (50)	1970-01-01 to 2021-01-31	relevance
"climate alarmism"	1 (50)	1970-01-01 to 2021-01-31	relevance

Table 2: List of Keywords indicating a Discussions of Interest

Keywords
youtube, yt, google, wikipedia, wiki, disclaimer, censorship, censoring, censored, restriction, restricted, demonitizing, demonitized, banner, warning, label, ban, removed, deleted, shadow, shadowbanning, banning, ban, banned, algorithm

added to a new data set. If a specific comment is added, we also check all replies to that comment (if present) and review the top ten most similar comments to make sure we are not missing on interesting information due to improperly clustering.
In the end, a data set for each video is created that contains only the comments of interest and give us the possibility to measure the proportion of comments actually dealing with platform interventions in the respective comment sections to answer the first part of the research question.

The following analysis is guided by the concept of the grounded theory [5]. The methodology involves the construction of hypotheses and theories through the collection and analysis of data. Ideas and concepts become apparent by reviewing the collected data. From the beginning of the research process, the researchers code the data, compare data and codes, and identify analytic leads and tentative categories to develop through further data collection [3]. In our case the videos were consecutively scanned and comments were marked, that can be summarized to arguments and positions in the debate. The process stops when the additional gathering and analysis of data no longer contribute to the understanding of the debate and does not yield to the formulation of new arguments.
Caused by the fact that a qualitative analysis always include personal bias, we decided to implement a concept where each analyzed video need to be qualitatively analyzed by at least two group members.

5 RESULTS

In this section, we focus on the empirical data in detail including the selection process of the analyzed videos, describing and summarizing the content of the videos, the number of the comments on platform interventions of each video, and the analysis of the post-video discussion.

5.1 Video Selection

We began the examination of comments with a set of five videos before we implemented the automatic pipelines for selecting videos. The selection of the first five videos was based on the number of the extracted comments that contain words from a preliminary keywords list (a sub-group of the final keyword list). At this level, the grounded theory enriched the pipeline by analyzing the comments consecutively, which led to the addition of keywords determining interesting comments.
The automatic pipeline determined twenty candidate videos. It is important to mention that not all of the candidates actually showed many comments on platform interventions although specific keywords were present. We selected videos that have a significant number of interesting comments in the light of our research questions. Consequently, we chose videos where more than 1% of the comments are related to our topic of interest.
The videos were all uploaded by different channels and the category labels of the videos were all among the categories news and politics, nonprofits and activism, people and blogs, science and technology, and education. Due to time constraints, the number of videos analyze qualitatively needed to be capped. The entire list of analyzed videos is given in Table 3.

5.2 Video Content

To get an understanding of the context, in which the users are arguing, the next section is a small summary regarding the content of the analyzed videos. There is definitely a huge variation in the content itself and even more so in the format of presentation. The conducted videos include documentaries, a TED talk, a CNN report as well as so-called infotainment videos. The presentation methods vary and range from interview scenarios to news presentations and graphic illustrations combined with a commentary background voice. In the following, we will describe three videos from which a significant proportion of the arguments described in a later section are taken from.
The first of the three videos is titled *"The Great Global Warming Swindle- FULL Documentary - Debunking Climate Change Hysteria"*. It is a British documentary which originally premiered in the UK television on British Channel 4 in 2007. It argues against the "virtually unchallenged consensus that global warming is man-made" and present changes in the radiation from the sun as the chief cause of climate change. According to the documentary, humans do have an effect on climate, but it’s infinitesimally small compared with the vast natural forces which are constantly pushing global temperatures. A group of scientists brought together by documentary maker Martin Durkin argued that everything one has ever been told about global warming is probably untrue and that it is the darkest chapter in the history of mankind. In their eyes global warming has only become a story of huge political significance, where environmental activists use scare tactics to further their cause. Scientists add credence to secure billions of dollars in research money and politicians and media are happy to play along. "Nobody dares speak

Table 3: Analyzed Videos

Video title	Video ID	Channel	Published	Category
How Well Do Scientists Understand Global Warming? | Climate Change on America Uncovered	xtBxI_ydba4	America Uncovered	2019-03-14	News & Politics
The Biggest Lie About Climate Change	TbW_1MtC2So	AsapSCIENCE	2019-03-14	Education
Neil deGrasse Tyson scolds cherry picking climate science	y1MZ8U8C9c8	CNN	2017-09-17	News & Politics
The Great Global Warming Swindle - FULL Documentary - Debunking Climate Change Hysteria	HHR9uJjrtAY	Indicrat	2018-08-19	News & Politics
The Great Global Warming Swindle	9gpEzUqCN7M	The Center for Investigative Reporting	2012-05-03	Entertainment
Why climate change is about money | Kashmala Kakakhel | TEDxIslamabadWomen	dQLLXyFvQOE	TEDx Talks	2019-01-14	Nonprofits & Activism
Global warming: why you should not worry	pwvVephTIHU	The Boston Globe	2010-05-14	News & Politics
What's REALLY Warming the Earth?	hphdsLcSTYQ	It's Okay To Be Smart	2016-08-29	Education
Climate Alarmism Isn't Rational	tUR0LrSadkg	PragerU	2019-09-06	Education
Debunking Tony Heller's 'Gift to Climate Alarmists'	XjPkclkZh6o	Mallen Baker	2019-10-01	People & Blogs

against it for risk of being unpopular, losing funds and jeopardizing careers."

Another video, whose comments are heavily reflected in the later stated arguments in the discussion on platform interventions is named *"How Well Do Scientists Understand Global Warming? | Climate Change on America Uncovered"*. The concept of the show is that subscribers were asked beforehand what they wanted to know about climate change and the moderator (and his team) explored how well scientists "understand" the respective topic and how much consensus there is without taking an own position. Within the course of the report many arguments are made for the uncertainty as to whether climate change is actually man-made or not. However, the conclusion was mainly that many scientists merely state that there is not enough evidence to confirm that humans are the most significant factor for the global warming. According to the host the video should not even be controversial, but he states within the first 30 seconds of the video, that YouTube is probably trying to "preempt" it by linking the Wikipedia entry on global warming below the video. Since the Wikipedia label was added, a comprehensive discussion started in the comment section.

The third and last video described in this section has the title *"The Biggest Lie About Climate Change"* and presents a cartoon-like animated video which tries to explain the emergence of climate change denial produced by a channel called AsapSCIENCE. The protagonist clearly pointed out his belief in the existence of a man-made climate change and described how climate change became a political issue. Using the example of American oil companies, the miscommunication of research results and the resulting skepticism towards climate change are shown. Due to the clickbait-oriented title of the video, people with very different opinions watched and commented on this video. Therefore, the comment section provides a lot of interesting debates not only but also on platform interventions.

5.3 The Number of Comments on Platform Interventions

After applying the pipeline and the subsequent qualitative analysis, we are able to discern the approximate number of comments on platform interventions for each video, as shown in Table 4. Overall, the highest percentage of comments dealing with platform interventions of a single video is 90%. However, this video only holds a total of 40 comments and was marked with a wikipedia entry on global warming below the video, which was the center of the discussion in the comment section. The second highest proportion of comments of interest was 6%.

For two videos, whose number of the total comments surpassed 12,000 comments at the time of data extraction, the proportion of the comments on platform interventions does not exceed 3% of the total comments. Nonetheless, the 2% and 3% proportions of these two videos correspond to 246 and 383 comments of interest.

Excluding the video with only 40 comments, the average proportion of comments on platform interventions was 3% corresponding to an average number of 99 comments.

5.4 Analysis of post-video discussions

In primary themes, the observed debates can be summarized in the following points: how users interact with platform interventions, how users understand and justify the policy of moderation that YouTube applies, and which cases are required to be moderated. In addition, users discussed platform interventions with a focus on the credibility of the platform and the credibility of the presented information, and how platform interventions are consonant with the right to free speech. The last point, we will present under the title of critical thinking.

In this section, we will present the results of the qualitative analysis in six categories of coding to detect in more detail how each of adding an informative banner, removing videos and comments, and the YouTube search algorithm were discussed based on the previous themes. The topic itself (climate change) and how it was broached by the content of the individual videos have in some categories a powerful impact on how users evaluate the moderation steps taken. Whereas, in other categories, the previous themes were discussed regardless of the topic and the attitude of the commentators toward climate change whether they are "believers" or "deniers".
Preliminary remarks:

- In the following paragraphs, we supported the extracted results by examples of the comments (in italics), we added the interesting part of the sampled comments and kept the comments as they have been written by the users without correcting the typos and other formatting specifics.
- Most of the comments are shared between two categories and sometimes more. Therefore, we did not calculate the absolute number of the comments in each category rather we tried to mention how prominent each attitude is.
- A significant number of comments mentioned that the commentators managed to find the videos, and since the informative banner is noticeable under the videos, each of YouTube search results and adding the informative banner were the main incentive of the discussion of platform interventions.

5.4.1 ***Credibility.*** The discussion on the credibility in the comment sections of the selected videos on climate change depends in the first place on how users evaluate the credibility of the information and extended to debate the credibility of the platform and the expected role of social media in dealing with specific public issues. Since the credibility is one of the most important indicators that determine the relationship between the platform and the users,we will illuminate different contexts of how users discussed credibility in detail, how they related it to each of the platform interventions, and what its influence on the different attitudes.

Source credibility and message credibility
Comments in this category debate on adding an informative banner with a link to a Wikipedia article about global warming and banning videos as platform interventions. In fact, adding the banner is the most noticeable intervention by the users. Therefore, it is the first stimulus of the whole discussion that we are interested in. In this part, we will clarify different attitudes of users toward adding the banner and focus on source credibility namely how they argue the validity of Wikipedia as a source of information.

Wikipedia defines itself as a free-content, multilingual encyclopedia project operated by a non-profit organization, and it is one of the Webs most popular reference sites among internet users. In the comments section of the selected videos, some users stood up for using Wikipedia as a source *“nothing wrong with using wikipedia as a source.”* Whereas, others who suppose that Wikipedia is owned by Google related choosing the used resource to the authority of power and money in governing the world and manipulating the science. Furthermore, users mentioned that the credibility of Wikipedia is limited in dealing with sensitive topics like climate change. From this point of view, some comments show worries about the sources that could be used in the future when big news media corporation wants to pay google for adding banners and include their sources under videos: *“at least its wikipedia for now”.* This confusion between Wikipedia and Google opens the question on the credibility of the medium and it will be discussed later on.

On the other side, using Wikipedia as a source was objected to by most of the users who wondered if YouTube relies on the narratives in presenting the reality of climate change *“Is YouTube trying to drive a narrative?”* and used many arguments against Wikipedia, one of them is that Wikipedia is not admitted as a valid source in the universities *“Wikipedia is not a fact based platform. Stop Censorship !”.*

For banning videos, the debate focuses on the message credibility including the content of the videos, the quality of the information, the strength of the arguments, and the credibility of the cited sources where few users found that banning videos that argue against climate change could be reasonable and acceptable. Justifications for removing videos could be invalidity of their content in the light of new scientific discoveries or due to presenting misleading information that underestimate or deny climate change: *“they didn’t do it because the video is ’controversial’, they did it because the video is **wrong**”.*

Platform credibility
Our analysis uncovered that the users relate the platform interventions, especially the YouTube search algorithm and banning videos, to the credibility of the platform. The debate on the credibility of the platform varies based on how users define the platform and evaluate its functions. In this regard, we can differentiate three different user attitudes: users who admit the credibility of the platform, users who think that YouTube is a commercial organisation encouraging creators of the videos to make videos that people want to watch in order to increase its stocks, and users who see that YouTube is a biased platform trying to control users with a systematic agenda. Depending on the factors that identified medium credibility (dependency, interactivity and transparency), we present the essential points of this discussion in the following.

Since YouTube is a source of information for an increasing number of users, some comments emphasized the credibility of YouTube

Table 4: Proportion of Comments of Interest

Video ID	Total comments	Comments of interest	Comments of interest (%)
xtBxI_ydba4	2481	99	4%
TbW_1MtC2So	12,751	383	3%
y1MZ8U8C9c8	12,303	246	2%
HHR9uJjrtAY	249	15	6%
9gpEzUqCN7M	40	36	90%
dQLLXyFvQOE	363	9	2.5%
pwvVephTIHU	2300	49	2%
hphdsLcSTYQ	964	14	1.5%
tUR0LrSadkg	1240	60	4.8%
XjPkclkZh6o	1603	21	1.3%

and rejected any question on it as an explanation of the moderation *"YouTube is the Real platform for answers"*. For instance, adding the banner is considered a credible sign of conspiratorial content, as users described. This attitude shows the effect of the dependency on how users experience platform interventions.

On the other hand, banning videos and comments on specific topics like climate change, Covid-19, and vaccines and the strict policy of YouTube dealing with were considered as an indicator of the lack of credibility: *"Again, you lose all credibility by hiding."*. In order to promote lies and mislead the people *"They'll delete this comment again because liars fear the truth"*. The search algorithm of YouTube was also broached due to the difficulty of finding videos against the consensus about climate change among the search results even if the users use the exact title of the video or because giving a space for such videos.

Some users critiqued YouTube for popularization conspiratorial videos that suspect the reality of climate change or underestimate its consequences *"Youtube has become a breeding ground for conspiracy theorists"*. In this regard, a considerable number of the users shared the point of view that YouTube and Google manipulate the people by controlling the search results *"stop manipulating the search, YouTube"* and forcing the users to watch and believe what coincides with the a agenda of Google, the owner of YouTube *"people watch and hear what they are forced to"* From this point of view, the platform aims to present one side of the discussed issue, distort the reality, and introduce artificial realities that serve the goals of Google as a profit company *"Powerful institutions like google are working flat out to get us to substitute reality with augmented and artificial realities. It's here but it ain't real"*. Many commentators went further and doubted the number of views and likes which affects the position of the video in the search results and thus its visibility *"Don't believe the low number of views. Mere manipulation via crooked globalist censorship"* and it might be out the tendency of users to measure the quality of a video by its views.

5.4.2 *Politicization of the debate.*

The politicized nature of discourses on climate change is studied as a common phenomenon in social media analysis by many researchers [20], [23]. In the comments on platform interventions we analyzed, the discussion on platform credibility was at some point expanded to a more generally political one. This trend is mostly influenced not only by users' attitudes towards climate change but also by the political attitude of users. This attitude correlates climate change activism with a left-wing political affiliation and climate change skepticism with conservative and right-wing thinking.

In this category, users explain platform interventions with their assumption of the political leanings of YouTube, as well as the political inclination of the channel or the youtubers. They consider videos against the science behind climate change as right leaning videos.

Some commentators saw their expectation about the leftist bias of the moderation policy of YouTube confirmed, especially in regards to the banning of certain videos *"Right leaning videos are demobilized or banned far more often than leftist"*. Platform interventions in this debate were evaluated as actions by YouTube as a leftist organisation, i.e.*"You tube is a leftist organization "*.Thus, YouTube is not viewed as a credible platform as it is perceived as trying to suppress any opinion controversial to its own, *"The only way the left can win a debate is to silence their opposition."*. YouTube's search results, namely pushing videos against climate change down, were deemed as a tool to prevent the popularity of right-wing activities. A remarkable number of the comments asserted a leftist tendency of YouTube, Facebook, Twitter, and other social media platforms, which they describe as a propaganda machines. Right along with this thinking, they seem to understand Wikipedia as a leftist source. This becomes especially relevant in the case of adding the informative banner with a link to Wikipedia article on climate change to relevant videos *"Google Youtube propoganda machine has put a WIkipedia that is left biased."*. On the other hand, users who we would locate on the left of the political spectrum, generally, seem to be more in favor of removing videos denying climate change, which they would consider right-wing videos. Additionally, they condemn that YouTube gives voices that paid by right wings a chance to be heard *"well paid YouTube propagandists for the Fascists who prefer the term alt-right"*. Comparing the number of comments that support this view to the number of comments for the previous one, we can say that the attitude of explaining platform interventions with leftist biases of the platform is the salient attitude in discussions of the political reasons behind platform interventions.

5.4.3 *Platform interventions and critical thinking.*

Some users frequently criticize YouTube for adopting only one point of view and shutting down different opinions or different interpretations of the data on climate change. As we discussed in previous sections, this was debated by users in different contexts and related to the credibility of the platform as well as the political tendency of it. In this argument category, the commentators discussed the legitimacy of platform interventions, how they are contrary to the freedom of speech, and how far the platform is responsible for the introduced content. Thereby, they are focusing on perceived transparency of the platform when it comes to its actions and motivations. Furthermore, this discourse illustrates how users evaluate the role of YouTube and social media generally in influencing public opinion toward specific issues.

On one side of the debate, there are users who seem to value the transparency of the platform and show less dependency on YouTube as a source of information, They seem less interested in the interactivity demanded when expecting YouTube to be an open platform for sharing ideas, opinions, and information freely. Therefore they rejected removing any video even if its content is invalid or terrible *"So why would YouTube remove it? It's opinion. It's allowed by freedom of speech"* and the same for hiding comments. Furthermore, they confirmed that the legitimacy of platform is based on the freedom of speech *"This platform gets special legal privileges from the American people for being an open forum - free speech and all that stuff."*. As long as this attitude condemns the concept of the current moderation policy, at least in dealing with climate change, we can only present how users related platform interventions to freedom of speech in the comment section of videos on climate change. We have no information about how these users would broach other cases such as the moderation of the content that includes, for instance, hate speech. In the same context, users also debate the credibility of the information that the videos present. They consider gaining knowledge mostly as an individual responsibility that can be objective by taking different views into account *"You have to make you own conclusion after considering both sides of the topic."*. They, then, continue to argue that giving the opportunity for different views to be discussed is the responsibility of a platform like YouTube.

On the other side, some users think that publishing videos and information that suspect the importance of climate change and human responsibility for it is a dangerous practice increases the threat of climate change and its consequences risk which have to be handled seriously *"YouTube must of hid this video. Is there a freedomtube?"*. In this regard, platform interventions are coherent with how users understand the social responsibility of social media platforms. Especially, when they view the platforms as one of the most important elements in guiding public attitude towards specific topics.

In this category, platform interventions were discussed generally according to how users define the role of social media platforms and what they expect from them. Based on the comments, we can mark a relation between the legitimacy of platform interventions and the dependency of users on the platform and consequently a relation between rejecting platform interventions and the expected transparency.

5.4.4 ***Users attempt to understand and explain the platform interventions***. An important point in the discussion of platform interventions is if users understand how the moderation policy of YouTube and algorithms, such as the search algorithm, work. User attitudes in this category are influenced by five factors: the attitude of the user towards the platform and its credibility, the ability of the user to evaluate the credibility of the information, including the content of the videos and the cited sources, the attitude of the user toward climate change, the relation between the users and YouTube as a platform, and how familiar the user is with the moderation policy of YouTube and its updates.

The analysis reveals a considerable number of comments where the user presents their thoughts regarding platform interventions and explains the moderation policy to other users. Thereby, the focus of the explanation is on how YouTube seems to be hiding comments, banning videos, and the YouTube search algorithm in general. Normal users, as well as youtubers with their own channel, share their experiences and expectations.

For some commentators, hate speech and violence are the only viable reasons for banning videos, which makes it unreasonable in the case of videos on a topic such as climate change, *"this video will be demonitized by youtube and labeled hate speech."*.

This misunderstanding of the moderation policy affects the attitude of the users toward the platform credibility. The doubt about the platform credibility, in return, leads to prejudice towards its actions.

Additionally, the users attempt to figure out how far intertwined the content of the video of interest is with content that would meet the restriction criteria of YouTube *"none of the restricted videos actually meet the community guidelines for restricted content."*.

One could argue that this reflects a more comprehensive understanding of YouTube's moderation policies and moderation in general. Regarding the YouTube search algorithm, some users think that the algorithm ensures that videos that use old resources or wrong information will not be higher up than in the search results or more prominently in the recommended videos than more valid videos *"they have a responsibility to ensure the correct information is given priority"*.

Judging from this perspective, the quality of the video content is the only factor that should determine the order of the video in the search results. The ordering would have to be controlled by a trustful site. In the comment section of the video "Debunking Tony Heller's 'Gift to Climate Alarmists' ", there is a long discussion on the search algorithm between users and a youtuber who explains the role of using SEO as an important option offered by YouTube for youtubers to ensure more visibility of the videos[1]. *"I aim to use good SEO with my videos within the mechanisms of YouTube, which includes adding appropriate tags for the my videos."*.

In addition, some users seem suspicious of both the number of likes and the visibility of the like-dislike ratio in the video. There is speculation in some comment sections if lack of visibility is a kind of platform interventions or even set by the channel itself *"YouTube has nothing to do with it. Likes/Dislikes visibility is set by the channel owner when the video is uploaded."*.

Many comments in this category agree that climate change deniers often tend to demonize YouTube and could not consider any

[1] SEO is the process of optimizing videos, playlists, and channel to rank high in YouTube's organic search results for a given search query.

option chosen by the video makers far away from the conspiratorial logic.

5.4.5 ***Practices against platform interventions.*** In this category, we do not focus on how the users discuss platform interventions but how their comments could be considered as practice against them. This attitude indicates how upset some of the users are about the moderation policies, especially hiding comments and banning videos. These users are actively trying to find solutions to keep sharing videos and ideas that they support. Among the comments we analyzed, we noticed that users consistently mention other videos that have already been banned. It seems that there are a number of controversial videos that are well known to users. These users might follow the video or its creator on other platforms, share their links, and invite each other to download videos that are expected to be banned in order to republish them later or elsewhere *"A good reason to download this to your computer. If YT this delete, we can re-publish it again! ;-)"*. Sharing links to specific videos existing on YouTube is used to facilitate finding these videos. This become relevant as the results of the YouTube search algorithm do not show them currently or might not show them anymore in the future.

As a practice to combat having comments hidden by the platform, some commentators attempt to find alternative ways and platforms to contact each other, such as Facebook, or forums to share videos and articles and argue on climate change without interference.

There are a number one comment typographical error we noted in the comments on platform interventions, as well as on YouTube and Google. When users critiqued the platform as a whole or tackle specific moderation measures, they might type "you tube" or "WIKIpedia". We can only speculate whether the typo is deliberate in order to manipulate the moderation algorithms. This behavior is seemingly more common among climate change deniers, which is congruent with higher degrees of expectations of being banned and censored among these users.

5.4.6 ***Other attitudes.*** As we have presented in previous sections, users discuss and explain platform interventions in different ways. This category of comments is dedicated to comments of users who responded to and mentioned platform interventions without their comment revealing further explanations that would clarify their attitudes.

Some users directly ask YouTube and a team of moderators to take action in the form of removing videos denying the science behind climate change or hiding comments of the deniers *"delete climate denier's comments"*.

On the other hand, the majority of comments we analyzed that explicitly mention platform interventions show that users do not seem to be satisfied with platform interventions. In these comments, users complain about the moderation and express their discontent about adding the informative banner *"What is with the Wikipedia link"* and the results of the search algorithm. According to the comments we analyzed, hiding possible search results is considered the most provocative intervention on the platform's part. Users complain also about showing such videos further down in the results and considering each question about climate change as conspiratorial *"Search for "Climate change is a hoax" on Youtube. No matter how hard you try, most of the videos are arguing against it"*. This includes also the use of ironic comments on the taken or expected interventions *"lol Actually at the moment its just pushed down"* and *"Brilliant, YT will delete this one day!"*. Unlike most previous attitudes we discussed, some users accept the platform interventions and receive them well. For instance, they seem to agree with adding the informative banner and do not grumble about it *"I love that YouTube keeps putting that banner up on any video about climate change"*. These users seem to find that banning videos and comments which argue against climate change or underestimate it is acceptable and necessary. The comments also mention that some youtubers claim to be banned to increase their popularity and bring attention to them. In our case study, this attitude is not as salient and more relevant to "believers" than the "deniers" of climate change.

Overall, since we are interested only in videos about climate change, we did not study the general influence of the video content or video topic on the debate on platform interventions. In other words, the same users that support banning and restricting information about climate change denial could have different attitudes toward platform interventions regarding other topics.

6 DISCUSSION

This section discusses our findings in more detail. The structure is as follows: In the first subsection, we contextualize the impact of a first comment about platform interventions in relation to the overall proportion of comments concerned with platform interventions. In the following section, we discuss how users refer each other to other platforms to re-upload deleted content or simply express themselves without fear of restrictions. We discuss how comments can be read as misunderstanding YouTube as a whole or merely the relevant guidelines and how users can be categorized in their attitude towards platform interventions. Subsequently, we consider the impact of the makeup of the target audience of our videos on the comment section.

6.1 Effect of initial Comments

Regarding the number of comments that actually refer to platform interventions in the comment section of a single video, a simple phenomenon can often be observed: A number of initial arguments of a video dictate the course of the following discussion. This observation becomes especially critical for videos with a lower number of views and an associated smaller number of comments. If one of the first comments expresses a specific opinion, many users take up this statement again and the comments on a certain topic pile up. On one hand, this creates replies to the original comments. On the other hand, there will be new comments that bring up own statements about the same circumstance, which potentially would not have been posted without the original comment that set the thematic direction for new posts. The primary statement got the ball rolling and the replies and follow-up comments are the results. If, for example, the initial comments of a video that addresses a certain topic (e.g. global warming) mentions the YouTube search algorithm, it can happen that the entire subsequent discussion only focuses on that the search algorithm instead of the actual video topic. We observe this exact scenario in one of the videos we analyze. The video in question only has 40 comments. However, it shows a 90 second excerpt of the *"The Great Global Warming Swindle"* documentary, in which many climate-related statements are made. Almost the entire comment section (36 of 40 comments) argues about the search algorithm and the possibility to find this video via the YouTube search function.

6.2 Referral to other Platforms

In the wake of a crackdown on extremist content, conversation in the YouTube sections, we analyzed, turned to strategies of saving and transferring existing videos, and moving the community elsewhere.

The debate about moving content mostly starts in the replies to one initial comment stating an expected removal of the video. The debate about moving whole communities elsewhere is similarly most often prompted by one initial comment. Users replying will either voice their agreement to the initial sentiment or note how the video is still up. Users claim they have already saved the content in the form of downloading the video and are ready to republish or transfer the video to another platform. To a comment stating *"Brilliant, YT will delete this one day!"* one user replied *"A good reason to download this to your computer. If YT this delete, we can re-publish it again! ;-)"*, while another simply said *"Been here for a year"*. Platforms like Parler gained traction and cracked under the increase in traffic stemming from the exodus of YouTube and other social media sites. Many conservative figureheads promoted alternatives to sites like YouTube by advocating like-minded users to come on to Parler, which promises its users to be able to "speak freely and express yourself openly, without fear of being 'deplatformed' for your views" [1, 4]. This mission statement echoes sentiments recorded by Myers West when interviewing users of several platforms who experienced being moderated [16]. With this promise, Parler became a haven for people having their accounts suspended from more traditional platforms for violating their terms of service. The hope of users flocking to the platform is that the kinds of debates and videos banned on YouTube would find a new home there. The same sentiments of being censored and restrained, Myers West reported on in her research, drive YouTube users fearing moderation into the hands of alternative platforms. However, following Trump's ban from YouTube, Twitter, Facebook, and other platforms, Parler itself became "deplatformed" from the Apple Store and Google Play Store [1]. Yet, even before the rise of alternative platforms even more intensely "free speech"-oriented than Twitter, links to YouTube videos not accessible through the results of simple search bar queries anymore were shared on other platforms, such as dedicated Reddit threads. On one hand, the relevant communities to those users speculating about the removal of a YouTube video are already elsewhere. On the other hand, even users strongly believing in climate change being a hoax and the discourse around it being a sign of some larger conspiracy against conservative thinking are not giving up on YouTube easily.

6.3 Attitudes towards YouTube as a platform

One sentiment echoing through many of the comments is the attitude of users towards YouTube being somewhat angry or disgruntled. There seem to be certain expectations towards YouTube as a platform. Reactions to bans and banners reveal an interesting phenomenon.

YouTube is and remains a commercial enterprise. However, some users seems to view the platform as a civic space; or at least one in which the same rules should apply. The expectation is that the ability to exercise "free speech" should be protected, no "censorship" should be allowed on the platform and if there is to be moderation it should affect all political affiliations equally. The videos we analyzed are mostly located on the conservative, right-wing side of the political spectrum. Comments noting an upcoming ban of a video seem to position YouTube "on the left" and understand deleting videos as an infringement against their right to self-expression or at least as part of an implicit agenda. To give just two examples: *"Well YouTube gives me the "accepted" definition of global warming under the video. They are already on the case"*, *"They don't want smart solutions. They want socialism."*. However, as a private entity, YouTube does not stand on the same legal ground as any other public space. The right to free speech is not applicable in the way that users seem to expect it to. The question that remains and is out of scope for this project is if it is an intentional misrepresentation of the legal situation or a sign of a lack of media literacy that drives people to express this opinion.

On a similar note, one could argue that the expressed expectation that videos will get banned stems from a misunderstanding of the applicable guidelines. There are several guidelines that apply on a number of different levels. When signing up for a platform, users have to agree to the terms of service. Additionally, there are community guidelines regulating the platform. Both terms of service and community guidelines are subject to change. They might not be phrased in a way accessible to all users or interpreted by all users the same. Especially, the terms of service are usually extensive and worded as a protection against litigation. These two comments would be exemplary for this phenomenon:
"Youtube allows political videos or religious ones or propaganda, but WE the users can say "I don't want to see them" [...] Your videos can be totally within community guidelines but still, be full of stuff that people don't want to see."
" @Augustus Maximus yeah, none of the restricted videos actually meet the community guidelines for restricted content.".
Both comments can be understood as responses to an expressed uncertainty about YouTube guidelines. Again, it is debatable if it is conscious misrepresentation of rules or if an inclusive, collective update of community guidelines would somewhat alleviate the situation.

6.4 How credibility is affected by users attitudes toward YouTube

Our analysis of the comment shows that different attitudes of users toward platform interventions are strongly influenced by their attitudes toward the platform's credibility. In this regard, we can differentiate three groups of users: users considering YouTube a credible platform and showing a tendency to accept and justify its actions including the platform interventions, users who are apprehensive of the platform's credibility but focus on discussing moderation efforts individually and independently of the platform, and finally users who do not believe in the credibility of the platform for different reasons and tend to reject its interventions regardless of the specific case that required to be moderated. This would suggest a kind of positive correlation between the evaluation of the platform's credibility and the evaluation of platform interventions and raises the question of how the users evaluate the credibility of the platform.

The comments show that users' attitudes are affected by their definition of the platform and social media generally, their function, and if the platform fulfills their expectations. YouTube, as the subject of our study, is merited for its diverse, rich content not only for entertainment and news but also for education. An increasing number of users consider YouTube a trustful source of information [15]. How users perceive the quality of the information or the cited sources strongly impacts how they rate the credibility of the message. Generally, how familiar users are with the YouTube ecosystem guides whether they rate a video as credibly if they do not have prior knowledge on the topic. Our assumption is that not all users have the same experience in assessing the quality of the source, the presented information, and the role of YouTube as the disseminating platform.

On the other side, users who define YouTube as an open space for sharing and expressing ideas place transparency on top of their priorities. In this sense, they might seem more objective in evaluating the credibility of the information, the credibility of YouTube and, consequently, the applicable moderation policy. Generally speaking, users evaluate information credibility based on informational factors such as the presented content, the strength of the arguments, and the quality of the resources. This requires diligent examination of the video and relates strongly to previous knowledge regarding the topic of interest and a general ability to assess the quality source material. Information credibility, on the other hand, is evaluated depending on information-irrelevant factors such as interactivity which might require less cognitive work. In other words, the information receiver's attitude towards the information can be affected by the ability and the motivation of the receivers [7], [11].

The attitude of users generally rejecting moderation is partly related to the worry about being manipulated by YouTube as a commercial company. This tension in the relation between the user and YouTube is sometimes linked to political considerations. As we have seen, some users think of YouTube as a leftist organization that tries to increase the popularity of the leftist ideas. In this context, the ambiguity on the exact goals of YouTube and how it works on accomplishing them affect the ways users consider moderation. Some users reject and condemn YouTube policies in order to protect themselves and their thoughts.

6.5 Consideration of the Audience

At this point, we deem it important to mention that our findings and interpretations must be treated with consideration of the general audience of the videos we selected. We deliberately chose the topic of climate change skepticism for our study due to a primary assumption: We consider a research field, in which there is a dominant scientific thesis, but also a noticeable amount of people who disagree as a great source of data. Especially, when it comes to investigating opinions on a third-party interfering with discussions on this controversial topic in the form of platform interventions. Indeed, the topic we chose has been a rich source of data. However, we do not claim to be able to infer the general opinion on platform interventions of the majority of users by investigating only a single field of discussion, such as climate change skepticism. Some conclusions may definitely hold universally, but individual controversies have always different characterizations and affect different communities. These communities might, then, differ in terms of personal perspectives of members and argumentation patterns, which in return trigger different responses from the platforms hosting them.

Recent noticeable changes in global climate and movements like "Fridays for Future" make the topic of climate change a field of broad social interest. Both supporters and deniers of the dominant opinion that global warming is mainly caused by human activities are present in the comment section. This may spark a more extensive and in some cases much more aggressive debate because people of fundamentally different opinions and belief systems argue with each other. If, for example, a similar study would focus on a conspiracy theory related niche without great social interest, the views represented in the YouTube's comment section regarding platform interventions might look completely different. Non-profits and other media have long speculated about the potential of radicalization through YouTube. A more obscure topic or the same video on another platform might make for completely attitudes [10]

In principle, the results of this project are very domain-specific and need to be interpreted with consideration of the target audience.

7 LIMITATIONS

In this project, we employed our experiences as data scientists in studying how YouTube platform interventions are discussed by the users in videos about climate change skepticism. The limitations of this work can be summarized in three important points.

The project would have profited from more team members in two ways. Conducting qualitative analyses is a time-consuming undertaking. All three of us are taking more courses on the side and cannot spend the time on this project that would be allotted to a full-time researcher. Additionally, our relative lack of experience in qualitative research required more deliberation of us regarding how to evaluate and analyze our source material and contextualize the findings in current research on the topic. As is, the time of the project was not sufficient to do a more comprehensive qualitative analysis and go further into detail with our research questions. This includes also collecting and analyzing more empirical data until the findings converge and do not cause significant changes in the results.

Secondly, there is the challenge of finding videos about climate change skepticism. This limitation was equally part of the discussion as it is a direct result of the moderation policy of YouTube. Since platform interventions were rarely talked about in the remaining videos, as we have seen in the results section, it was hard to find videos that contain a significant number of comments related to our chosen topic. Despite the important role of the automated pipeline for selecting potentially interesting videos, not all of the candidate videos actually provided comments on platform interventions and interesting statements regarding research questions.

Finally, specifying the clusters of interest by the pipelines depends on the frequency of the words contained in the keywords list. In this regard, we expect that expanding the keywords list or improving it to avoid unrelated contexts could lead to more accurate results. Furthermore, there is the general problem of the polarity of words. The same words can have contrasting meanings in different contexts.

8 CONCLUSION

This seminar project aims to answer the question of how different types of platform interventions are discussed in videos about climate change skepticism on YouTube. To answer the question of how platform interventions are discussed in the comment section of climate change skepticism, we analyzed the relevant comment sections of ten videos.

We figured out that comments discussing platform interventions make up a varying degree of the comment sections. However, a large proportion of relevant comment are replies to one initial comment starting the discussion. The larger trend of moving controversial communities to other platforms is noticeable here, as well. When users comment on how the video is in danger of being removed, others are quick to point out that they have already saved the video and are ready to move elsewhere. The change in the YouTube algorithm is also heavily featured in the comment section. Users bemoan changes but it becomes clear that if users know where to look the same kind of content that is banned from more traditional platforms, such as YouTube, is still accessible. Further, there seems to be a disconnect in how users understand YouTube as a platform and real legal imperatives governing social media sites. When it comes to evaluating how reasonable platform interventions are, we identify three groups of users: users accepting platform interventions and considering them justified, apprehensive users who focus on individual moderation efforts and view them independently of the platform, and finally users reject all forms of moderation as an infringement on their personal freedoms. In which category a user falls seems to be motivated by a number of factors, such as whether users consider YouTube an inherently political site, their levels of media literacy, and how much they equate moderation with censorship. Overall, the topic we chose and the related target audience of its videos play a significant role in what the comment sections entail.

9 OUTLOOK

As previously described there is a clear bias by the target audience in our findings. That bias can never be eliminated but decreased by investigating a diverse collection of controversies and conspiracies and how they argue on content moderation and platform interventions in general. This would be the case if a study tries to work out universally applicable opinions, but there is also a great benefit in highlighting specific opinions only in a small field of discussion as it was done in our study.

Regardless of whether the scope should be more universal or it is aimed to focus on a single field of interest, a more detailed analysis is definitely associated with the evaluation of large amounts of data. Here, the performance of a pure qualitative analysis reaches the limits of feasibility due to the enormous amount of time that need to be invested. In the nowadays fast evolving field of natural language processing there are already countless approaches for an automated sentiment analysis with differing complexities and capabilities. The automation of the sentiment analysis would lead to a simplified elaboration of arguments and offers the possibility to more or less effortless repeat the same analysis in constant steps over time to monitor trends and opinion shifts. Since YouTube is constantly implementing new mechanics to prevent misinformation and manage the handling of conspiracy theories, the over-time observation could provide very interesting findings to investigate the reactions on new intervention strategies.

Since YouTube's prevention methods and policies can never be ahead of the current state of knowledge in the science, it could also be very interesting to analyze cases where platform interventions were applied that had to be revoked in retrospect. Until something is proven in science there is always a possibility that it turns out to be true or false. Given that YouTube's intervention concepts are far away from perfect and that the state of knowledge in science is evolving very fast, the decision of content moderation by mistake will probably be debated very intensively and provide an interesting source of data for an exciting analysis.

REFERENCES

[1] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. 2021. An Early Look at the Parler Online Social Network. *arXiv preprint arXiv:2101.03820* (2021).

[2] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Céspedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175* (2018).

[3] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded theory. *The Blackwell encyclopedia of sociology* (2007).

[4] Luciano Floridi. 2021. Trump, Parler, and Regulating the Infosphere as Our Commons. *Philosophy & Technology* (2021), 1–5.

[5] Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research.* Routledge.

[6] James Grimmelmann. 2015. The virtues of moderation. *Yale JL & Tech.* 17 (2015), 42.

[7] Simon David Hirsbrunner. 2021. Negotiating the Data Deluge on YouTube: practices of knowledge appropriation and articulated ambiguity around visual scenarios of sea-level rise futures. *Frontiers in Communication* 6 (2021), 11.

[8] Shan Jiang, Ronald E Robertson, and Christo Wilson. 2019. Bias misperceived: The role of partisanship and misinformation in youtube comment moderation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 278–289.

[9] Nicolle Lamerichs. 2020. User Tactics and Algorithms: A Digital Humanities Approach to YouTube and Tumblr. In *Understanding Media and Society in the Age of Digitalisation.* Springer, 35–54.

[10] Mark Ledwich and Anna Zaitsev. 2019. Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *arXiv preprint arXiv:1912.11211* (2019).

[11] Ruohan Li and Ayoung Suh. 2015. Factors influencing information credibility on social media platforms: Evidence from Facebook pages. *Procedia computer science* 72 (2015), 314–328.

[12] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[13] D Harrison McKnight and Charles J Kacmar. 2007. Factors and effects of information credibility. In *Proceedings of the ninth international conference on Electronic commerce.* 423–432.

[14] Erwan Le Merrer, Benoit Morgan, and Gilles Trédan. 2020. Setting the Record Straighter on Shadow Banning. *arXiv preprint arXiv:2012.05101* (2020).

[15] Amir Michalovich and Arnon Hershkovitz. 2020. Assessing YouTube science news' credibility: The impact of web-search on the role of video, source, and user attributes. *Public Understanding of Science* 29, 4 (2020), 376–391.

[16] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383.

[17] Jacob Perkins. 2014. *Python 3 text processing with NLTK 3 cookbook.* Packt Publishing Ltd.

[18] Bernhard Rieder. 2015. YouTube data tools. *Computer software. Vers* 1, 5 (2015).

[19] Bernhard Rieder, Ariadna Matamoros-Fernández, and Òscar Coromina. 2018. From ranking algorithms to 'ranking cultures' Investigating the modulation of visibility in YouTube search results. *Convergence* 24, 1 (2018), 50–68.

[20] Matthew A Shapiro and Han Woo Park. 2018. Climate change and YouTube: Deliberation potential in post-video discussions. *Environmental Communication* 12, 1 (2018), 115–131.

[21] Qinlan Shen, Michael Yoder, Yohan Jo, and Carolyn Rose. 2018. Perceptions of censorship and moderation bias in political debate forums. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[22] Tan Thongtan and Tanasanee Phienthrakul. 2019. Sentiment classification using document embeddings trained with cosine similarity. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 407–414.

[23] Julie Uldam and Tina Askanius. 2013. Online civic cultures? Debating climate change activism on YouTube. *International Journal of Communication* 7 (2013), 20.