

Mandate to overblock? Understanding the impact of the European Union's Article 17 on copyright content moderation on YouTube

Daria Dergacheva¹  | Christian Katzenbach^{1,2} 

¹Platform Governance, Media and Technology Lab, Center for Media, Information and Communication Research (ZeMKI), University of Bremen, Bremen, Germany

²Humboldt Institute for Internet and Society, Berlin, Germany

Correspondence

Daria Dergacheva
Email: dergache@uni-bremen.de

Funding information

Horizon 2020 Framework Programme,
Grant/Award Number: 870626

Abstract

This article presents the results of a study measuring possible overblocking due to copyright moderation and changes in the diversity of cultural products supply on YouTube in two European Union (EU) member states comparable in size and population, Germany and France. Both have adopted Article 17 of the EU Copyright in the Digital Single Market (CDSM) Directive (CDSMD) but Germany was 5 months ahead of France in changing its legislative regime for large social media platforms and copyright content moderation. The article assesses how content takedowns that were likely connected to copyright differed in these two countries and how this might have been influenced by the greater copyright moderation linked to the implementation of Article 17 of the CDSMD. Furthermore, the cultural supply diversity according to genre on YouTube was examined by applying Stirling's model of diversity and using a dual-concept diversity index. To the best of the authors' knowledge, this is the first attempt to measure predicted changes to overblocking and possible changes in cultural diversity on YouTube's infrastructure since the new copyright rules came into force in the EU. The findings show that during the period examined, 2019–2022, significant differences were identified between Germany and France in terms of the takedowns of videos from categories prone to copyright moderation. In addition, the content-level supply of cultural products diversity, measured by genre (channel categories), number of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. © 2023 The Authors. *Policy & Internet* published by Wiley Periodicals LLC on behalf of Policy Studies Organization.

videos, and subscribers, decreased in YouTube's available content in both countries but more so in Germany than in France. The study also makes a methodological contribution to the field of content moderation by reusing previously collected data to explore YouTube's infrastructure and cultural diversity within the framework of media economics. Platforms' content moderation practices and the potential impacts they have on cultural diversity are notoriously difficult to study, and this research establishes new methodological frameworks for further exploration.

KEYWORDS

Article 17, copyright content moderation, Copyright in the Digital Single Market Directive, copyright policy, cultural production, data access, European Union, platform governance

INTRODUCTION

The spring of 2019 saw thousands of people marching through several European Union (EU) capitals (Martin, 2019, March 2) to protest against the future EU Copyright Directive. Critics were concerned that a new liability regime would force social media platforms to automatically filter and remove content, and social media platforms would feature less diversity and more user censorship in the future. In an open letter prominent Internet founding figures and spokespeople warned that this legislation “would mandate Internet platforms to embed an automated infrastructure for monitoring and censorship deep into their network” (O'Brien & Malcolm, 2018, June 12).

The move to pass the EU Copyright in the Digital Single Market (CDSM) Directive (CDSMD) (2019/790) occurred in a context in which social media platforms were clearly becoming key players in contemporary societies (van Dijck et al., 2018) and artificial intelligence (AI) technologies were increasingly being presented as solutions to major societal problems (Katzenbach, 2021). Under growing public and political pressure, social media platforms have massively expanded their efforts to monitor and moderate content on their sites. Platforms have made considerable investments to rapidly expand their teams of content moderators (Roberts, 2019) and introduce algorithmic systems to automatically govern contested content (Gorwa et al., 2020).

The CSDM Directive is highly relevant to both the future role of platforms as intermediaries, as well as their impact on cultural diversity and access to culture. YouTube's latest Transparency Report from December 2021 showed that in the first half of 2021, YouTube needed to reverse more than 2.2 million of its content removals, based on user disputes and appeals. In other words, YouTube's automated regulatory Content ID system had generated at least 2.2 million unjustified copyright actions against its users on behalf of rightsholders (Keller, 2021, December 9).

However, the number of unnoticed take-downs that are unequally unjustified may probably be even much higher. Algorithmic moderation systems deployed by private platforms remain opaque, if not completely nontransparent, and are often considered “black box[es]” (Pasquale, 2016). This opacity is multifaceted since it includes not only technical aspects, as Gray and Suzor (2020) noted, but also institutional and legal issues. Compared

to situations where the judiciary is involved in the decision-making process, the lack of accountability and disregard for fundamental rights on the part of platforms is further amplified (Jacques et al., 2018).

The CDSMD was adopted by the EU and came into force in June 2019. Countries have had 2 years to implement the CDSMD into national law, but almost 1 year after the deadline, the EU Commission issued a press release on May 19, 2022, saying that Belgium, Bulgaria, Cyprus, Denmark, Greece, France, Latvia, Poland, Portugal, Slovenia, Slovakia, Finland, and Sweden had not yet notified the Commission on changes to their national legislation (Press Release, EU Commission, 2022, May 19).

Key controversies centered on Article 17 (1) which introduces a new liability regime for platforms that was thought to endanger free speech. The Article states that in the absence of authorization from the copyright holder, online content-sharing service providers (OCSSP) must meet three conditions to be exempt from liability for copyright infringement. First, they must have made all reasonable efforts to obtain authorization from the copyright holder. Second, they need to have made all reasonable efforts to ensure the unavailability of the specific protected content for which the rightsholder has provided relevant and necessary information. Finally, once a rightsholder provides an OCSSP with a sufficiently substantiated notice, the provider must act expeditiously to disable access to that content and make all reasonable efforts to prevent future uploads. If the provider does not fulfill these requirements, they are liable for any violation of the rightsholders' exploitation rights. As researchers have stated, Article 17 of the CDSMD ushers in a profound transformation: previously, access to protected content was granted unless proven to be infringing, but now, any material identified by algorithms will be withdrawn from public circulation unless proven to be legitimate (Reda & Keller, 2021, June 4). This change represents a significant shift in the way we approach content moderation. As argued by Quintais (2022), for large platforms, Article 17 CDSMD gives platform operators a strong incentive to implement automated filtering tools to comply with the best efforts and obligations in Article 17(4) [of the] CDSMD and limit their liability. So does the CDSMD constitute a mandate to overblock in the practice of content moderation?

Against this background, this empirical study investigates the changes to and influences on access and cultural diversity on social media and streaming platforms, specifically YouTube, in the context of the CDSMD and its implementation in Germany and France.

In this study we seek to better understand empirically how CDSMD and its implementation affect in practice platforms' content moderation, and subsequently the availability and diversity of content. Specifically, we investigate:

RQ 1. *In countries where CDSMD 17 has been implemented earlier, do we observe any significant difference in content removal behavior compared to countries that had implemented CDSM later?*

RQ 2. *How might copyright content moderation have led to reduced diversity in the content supplied on YouTube in the Germany and France?*

The first part of the study presents general findings on copyright takedowns on YouTube in the two EU countries between 2019 and 2022. To obtain these findings, the researchers tested a subset of videos from the largest YouTube study (conducted by Rieder et al. in 2020) before assessing, first, whether they had been removed by YouTube and, second, how the removals were related to the countries in question, the video categories, and other predictors (such as likes and engagement). The second part of the study discusses changes in the supply of cultural production, measured by changes in the variety and balance of the channels (genres) in Germany and France. The aim was to identify any changes in content

supply diversity over the period and determine whether these changes varied when the sample countries were compared. Thus, the second part of the study answered research question 2.

UNDERSTANDING CONTENT MODERATION AND CULTURAL DIVERSITY

We approach these questions based on methodologies and frameworks from both content moderation research as well as cultural diversity research.

Mapping YouTube and copyright moderation

Copyright moderation has always been strongly connected to economic interests, and even before the introduction of formal policies, platforms were pressured to monitor and police the content on their sites (Gorwa et al., 2020). One of the most extensive studies of the effects of notice and takedown procedures, mainly those on Google's Web Search, was conducted by Urban et al. (2017). The quantitative aspect of the study revealed that based on the 6-month data set containing 108,331,663 requests, most copyright takedown requests were sent by the music (44% of requests), adult entertainment (28.1%), and movie/television (17%) industries.

YouTube's copyright content moderation system gives copyright holders considerable decision-making discretion (Kaye & Gray, 2021), while the large number of disputes, as yet unavailable as content-level data on the moderation and nature of actions undertaken by YouTube, indicates the possible ambiguity of the content-moderation processes. Similar issues apply with platforms like Instagram, for example, as discussed by Witt et al. (2019).

A large-scale study of YouTube channels was undertaken by Rieder et al. (2020). This was the first and, due to the growing inaccessibility of the platform's data, perhaps the last large-scale description of YouTube's media system. The researchers relied on a sample of over 36 million channels and videos and explored the platform's media system in three main directions: stratification and hierarchization in broadly quantitative terms; channel categories, their relationships, and their proportions; and channels according to country affiliation (Rieder et al., 2020). More specific studies on the copyright content moderation and takedown numbers on YouTube are scarce, and has found highly varying numbers: Gray and Suzor (2020) used a random sample of 76.7 million YouTube videos and analyzed the videos' metadata on content and removal information. While in this study only approximately 1% of all uploaded videos had been removed due to apparent copyright violations, an analysis by Erickson and Kretschmer (2018) of videos that are highly susceptible to copyright controversies, such as parodies, revealed with 15.5% a far higher percentage of takedowns that might be copyright-related.

This article builds on this scholarship by identifying categories most prone to copyright takedowns to evaluate takedown rates in the two EU countries under study.

Cultural diversity in cultural production research

Our study on the potential impacts of platform regulation and copyright content moderation on YouTube's infrastructure changes is grounded in the framework of cultural diversity, which encompasses diverse aspects of media such as source diversity and content diversity. This comprehensive and intricate concept has been extensively examined in

media studies, as demonstrated by previous research (Deacon et al., 2021; Voakes et al., 1996).

Diversity is a multifaceted concept that can be examined from different perspectives, including platforms, cultural content, and cultural producers, which frequently intersect and influence each other (Poell et al., 2021). Diversity among cultural producers (creators) or in actual video content is not in the scope of this research. In the existing literature, qualitative approaches have frequently been adopted to examine content and the representation of various groups on YouTube and other platforms (see, e.g., Bishop, 2019; Duguay et al., 2020; Lin & de Kloet, 2019; Phillips et al., 2022). However, the questions formulated in this study refer to the level of cultural content as shaped by the intermediary via genres (in the case of YouTube, this meant channel categories, and not to the content's producers or narratives diversity. In this study, we approach YouTube as an intermediary which "formats" cultural products (Nieborg & Poell, 2018; Poell et al., 2021; Siciliano, 2022), much like intermediaries in the publishing industry, film production or music define genres and roles (Silver et al., 2016).

According to Havens and Lotz (2016), regulatory decisions in traditional media markets aim to maintain both diversity of voices and localism within media industries. However, with the emergence of various platforms and the ingenuity with which cultural producers create new content genres, it is worth investigating whether these dominant platforms genuinely encourage unrestricted creativity and diversity, thus providing new avenues for citizenship expression.

In a recent study conducted by Hesmondhalgh et al. (2023), the relationship between the music industry, digital platforms, and cultural formation was explored through a historical lens. Their research revealed how infrastructure plays a crucial role in shaping culture. Specifically, they focused on the music industry to demonstrate how platforms have functioned as the primary medium through which the democratizing and liberating potentials of the Internet's partially commons-based openness were gradually diminished or eliminated altogether. Thus, the initial promise of platforms as a more open and inclusive infrastructure has ultimately led to the creation of an additional profit layer for corporations, while limiting diversity within the industry. Hesmondhalgh et al. concluded that the infrastructure provided by these platforms has indeed shaped culture, but instead of realizing the emancipatory possibilities of an open internet, these promises have been gradually eroded.

Previously, diversity in the supply of creative work that provides cultural goods and services has been a focus area in the media economics, sociology, and communication sciences fields. Researchers have examined cultural diversity in the movie industry (Lévy-Hartmann, 2011; Moreau & Peltier, 2004), TV networks (McDonald & Lin, 2004), recording companies (Ranaivoson, 2010), publishing (Benhamou & Peltier, 2007), and broadcasting (Farchy & Ranaivoson, 2011). The listed studies referred to genre variety and concentration, as well as number of copies sold, as the main indicators of diversity supply in these cultural industries. Jacques et al. (2018) has also applied a similar method of calculating changes in diversity supply according to video takedowns of parodies on YouTube. Thus, in the second part of our empirical research, our study explores how the changing platform interface and algorithms of YouTube's content moderation, including those that follow the CDSMD Article p17th policy on the copyright, influenced the diversity of cultural production visible on the platform at the genre levels.

Stirling model of cultural diversity

While studying cultural diversity in the supply of cultural goods and services, researchers have used the Stirling model of diversity (2007), which was derived from models in economics, ecology, and information theory. It consists of the measurement of three components: variety, balance, and disparity. According to the model, the greater the variety, balance, and disparity, the greater the diversity. Farchy and Ranaivoson (2011) pointed out

that to assess the diversity of, for example, the music industry (or any system), it must be first divided into different categories, such as geographical origin, title, and genre.

Other researchers have highlighted that it is not always possible to define such components as a disparity in cultural production. For example, Benhamou and Peltier (2007) stated that the question of measuring disparity remained unanswered since it must rely on certain assumptions about the distance between, for example, one geographical origin and another, or one movie genre and another. In the case of YouTube, disparity meant relying on assumptions about the distance between one category and another, while variety meant the exact number of categories (or genres) the platform has. Much like in previous studies, two components of diversity in cultural production measurement were considered in this study: balance and variety (Benhamou & Peltier, 2007; McDonald & Lin, 2004).

We build on the Stirling model of diversity to assess the cultural diversity of YouTube channels in two EU countries using two components of the model—variety and balance—and track how this changed between 2019 and 2022. To our knowledge, this is the first attempt to use this media economics and sociology framework in platform studies while focusing on the supply of cultural production through genres (channels and categories).

METHODOLOGY AND DATA COLLECTION

To understand, how CDSMD and its implementation affect practice platforms' content moderation, and subsequently the availability and diversity of content, we first estimate metrics on copyright takedowns on YouTube as a baseline and then compare the availability and diversity of a sample of YouTube Channels in two EU countries and jurisdiction with generally similar characteristics, yet significant difference in CDSMD implementation: Germany and France. Germany was among the few EU member states to meet the June 2021 deadline for the implementation of Article 17 of the CDSMD, which came into force there on August 1, 2021 (Brieske & Peukert, 2022). France implemented the CDSM (EU) 2019/790 Directive (CDSMD) on November 24, 2021, but did not inform the Commission of this until at least May 2022. As identified by Angelopoulos, European Commission, Directorate-General for Research and Innovation (2022) in an expert survey across 11 countries of the EU, including France and Germany, there substantial differences in the national implementations of the new copyright liability scheme for OCSSPs (“online content-sharing services providers”) which Article 17 of CDSMD implicates. With regard to France and Germany, her study highlights that France has ignored the suggestion of the Directive that authorizations granted to users to upload protected subject matter do extend to OCSSPs. This contradiction will most probably lead to an intervention by the Court of Justice of the European Union. In contrast, Germany has taken a proactive interpretation, establishing an approach that aims to safeguard freedom of speech with regard to Article 17 and to prevent systemic overblocking as feared by critics.

In this way, the comparison of the two countries is similar to a quasi-experiment in the context of the CDSMD, to investigate our research questions on significance of differences in content removal behavior between countries with different CDSMD timelines (RQ1), and on the impact of copyright content moderation on (reduced) diversity in the content supplied in those two countries (RQ2).

Assessment of copyright takedowns and comparison between Germany and France

For the investigation of differences in content removal behavior, we investigated and compared to data sets of a sample of EU-based video channels on YouTube. The first data

set was collected by Rieder et al. (2020) via the YouTube API v.3, well before CDSMD implementation. It constitutes of a sample of 4,000,000 videos from EU-based YouTube channels. Based on this, we have collected a second data set in 2022 after CDSMD implementation and filtered by countries (Germany and France). This second data set is a 2.09% subsample of the original data set, resulting in 83,676 videos.

These data sets allowed us to identify and compare the videos that had been blocked or deleted between June 2019 and June 2022. The data set includes metadata for blocked or deleted videos (from 2019). We then scraped pages of deleted videos on YouTube to identify the reason for deletion that YouTube provided at the time. Further, we used a random forest prediction model to assess each of the variables that were important for the take down where the reasons provided by YouTube were “copyright infringement complaint,” “unknown,” and “deletion of associated YouTube account.” As next step, we checked whether there was a significant difference between takedowns belonging to the categories prone to copyright claims (music, film, and entertainment) between Germany and France.

Comparing YouTube's infrastructural change in creative goods supply and demand in Germany and France

To collect data on the diversity of the creative goods supply on YouTube in Germany and France, the authors built the first breadth-depth YouTube crawler in Python by channel, country, and number of subscribers, which utilizes the YouTube API. As a point of entry, a random channel was used from the list of channels provided by Rieder et al. (2020). The crawler worked for 10 consecutive days on each country, going as far through the channel subscriptions as the YouTube API v3 quota allowed. As shown in Figure 1, channel

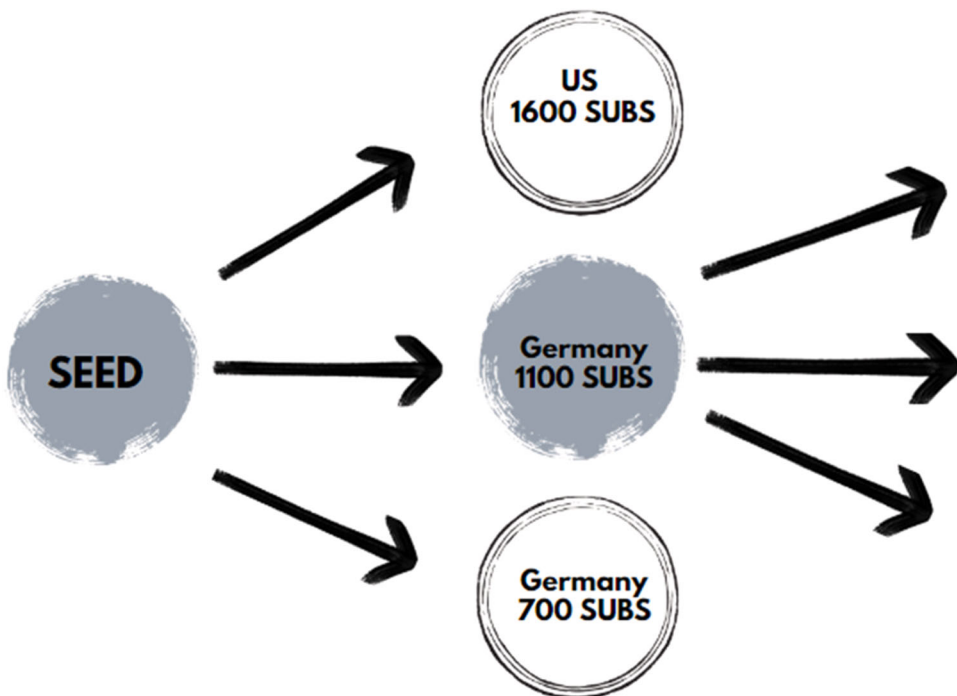


FIGURE 1 Principles of crawling YouTube for channel sampling.

subscriptions were crawled; when the crawler met a channel from the indicated country with over 1000 subscribers, it followed this path and collected the Channel ID in our data set, else it stopped (e.g., with channels from different countries or with fewer than 1000 subscribers)

Thus, by following the data collection methodology used by Rieder et al. (2020) but with a more limited capacity due to the reduced access to data when using the YouTube API v3, the authors built a random sample of channels with over 1000 subscribers from Germany and France.

In the next step, the channel sample gathered in 2022 was compared with the 2019 sample gathered by Rieder et al. In both samples (from Germany and France), at least 5% of the channels had 1000 or more subscribers. The authors then calculated and compared the changes in YouTube's infrastructure and cultural production supply in the two countries, based on the variety and distribution of the channel categories (genres) and the number of channels in each category. Much like in the movie or publishing industries, the world's largest streaming platform, YouTube, uses genres (categories for both channels and videos) in its infrastructure.

For measuring diversity, we used the framework suggested by Moreau and Peltier (2004) for measuring diversity in the movie industry. They assessed cultural diversity in the movie industry using the Stirling model of diversity (2007), relying on supplied diversity and two units of analysis: variety and balance. To study supply and consumed diversity through this prism of media economics and by regarding YouTube as an intermediary, the authors proposed to study YouTube in the same way. Thus, the two dimensions (variety and balance) and the two units of analysis (the channel category (or genre, which it actually is) and the number of channels belonging to the genre) would serve as indicators of cultural diversity supply.

As in the study by Moreau and Petlier of the movie industry and in the later study by Benhamou and Petlier of the French publishing industry (2007), these two units of analysis were used to assess the economic dimension of cultural diversity through a market concentration-related index, the Herfindahl–Hirschman Index (HHI). This index measures diversity allowing for variance in performance metrics. For the current study, this metric allows to measure the diversity of YouTube using genres (channel categories) and numbers of channels, indicating how diverse the content portfolio of the platform is in each country, at least in terms of genres that YouTube's channel creators tend to use. The HHI additionally enables us to use view counts, like counts, and comment counts as performance metrics in the prediction model to better assess diversity across countries and channel categories. More specifically, we build on Jacques et al. (2018) and their HHI-based concept “supplied diversity” that pertains to the range of content accessible to the public on a platform. As the authors highlight, although the HHI is commonly utilized in industrial economics to gauge market concentration among firms, it can also be employed to measure diversity when the subject being studied is classified into different types. With no variety but with complete concentration into a single variant, the index is equal to one. Where more variety and balance are available, the index is closer to zero.

RESULTS AND DISCUSSION

Assessing the role of copyright content moderation in takedowns

Underreporting of copyright-related takedowns in YouTube's metadata

Our investigation revealed that within our sample 6311 videos had been taken down (7.54%). This result was achieved by comparing the 2019 and 2022 data sets and noting

whether the metadata on these videos, while present in the original 2019 data set, was no longer there by 2022. To find more information about deleted videos, we subsequently scraped YouTube's actual pages for these videos. We then analyzed embedded information, specifically the `previewPlayabilityStatus` using the `htr` and `jsonlite` packages in the R programming language to determine potential causes of the deletions. The reasons provided by YouTube for video deletions are presented in Table 1.

These results are hard to interpret, though. While this data only directly attributes unavailability to copyright infringement in eight cases, there is reason to conclude that many of the remaining removals in the categories “unknown reasons” and “deletion of associated YouTube account” are just as well attached to copyright content moderation. First, previous scholarship (e.g., Erickson & Kretschmer, 2018; Jacques et al., 2018) treated all deleted videos from the parody sample as likely to have been removed in connection to copyright moderation (with one exception of “censorship”). Second, YouTube's enforcement process leads to structurally concealing copyright-related removals when looking at data on the content level. YouTube uses a system of copyright infringement “strikes,” which causes a channel being deleted after three infringements (YouTube Help, 2023). In consequence, videos of these channels are no longer available and they are earmarked as “Deletion of Associated YouTube Account”—although this mechanism is directly linked to copyright content moderation. Third, even the ContentID process (as YouTube's key copyright content moderation mechanism) does not result in clear flagging of removal due to copyright issues. In a first step, copyright owners (and perhaps the platforms themselves after adoption of Article 17 of the CDSMD) may choose to block a video in certain countries, monetize or track the video's statistics. None of these actions result in a “copyright strike” for the channel owners or are marked as copyright infringement complaints. If required, users may choose to remove claimed content or to appeal the claim within 7 days (YouTube Help, 2023), in order to avoid a strike. Removal by users results in labeling items with “Unknown reason...”. Only if creators wait for more than 7 days or appeal and get their appeal overrun, the platform removes the video and it gets earmarked as “copyright infringement complaint” (YouTube Help, 2023). From previous research, we know that the vast majority of users do comply with such requests and do not appeal (Kaye & Gray, 2021), so most of the videos identified by ContentID or receiving another form of takedown request and chosen to be deleted by the owners are most probably classified as “Unknown reason.”

In sum, this first assessment on the content-level clearly shows massive underreporting of copyright-related takedowns by YouTube in the embedded information of removed

TABLE 1 Sample of videos and YouTube's embedded information on availability (`previewPlayabilityStatus`).

	Number of videos	% of videos from sample
Videos in sample	83,676	
Videos unavailable	6311	7.54
Video unavailability due to privacy settings	2774	3.32
Removal due to complaint for false advertising	1	0.04
Unavailable without copyright relation	2775	3.32
Unavailability due to copyright infringement complaint	8	0.01
Deletion of associated YouTube account	383	0.46
Unknown reason	3145	3.76
Unavailable with potential copyright relation	3536	4.23

videos. As we have shown, the label “Unavailability due to Copyright Infringement Complaint” is not a good proxy for assessing effects of copyright content moderation. There is a strong indication that many of the videos in both the “unknown reasons” as well as the “account deleted” category are actually copyright-related.

Assessment of predictors for takedowns

To further assess the relation of copyright content moderation to removals for this underdefined content, we have developed additional statistical measures. For this, we have sought to identify metadata information that correlate highly with probable takedowns. In other words: We have statistically determined predictors for videos being taken down for copyright reasons. We used a random forest model, specifically the mean decrease accuracy plot, to assess each variable that was important for takedowns, with the reasons provided being “copyright infringement complaint,” “unknown reason,” and “deletion of associated YouTube account.” The model used variables derived from the video metadata provided by the YouTube API v3, as presented in Table 2.

Statistically, a mean decrease accuracy plot expresses how much accuracy a model loses by excluding each variable (Martinez-Taboada & Redondo, 2020). The more the accuracy suffers, the more important the variable is for a successful classification. In Figure 2, the variables are presented in descending order of importance.

As a result, category ID was revealed as the most important predictor of videos being taken down for an “unknown reason.” Witt et al. (2019) used a similar method of calculating the probability of removal for each category in Instagram's moderation of women's bodies. They also extrapolated the coded sample to a general population. Building on this result, we identify those categories that are most prone to copyright enforcement from the existing literature. In their large 2017 study of take-downs, Urban et al. (2017) found that music, adult entertainment, and movie/television were the types of content that received the most significant numbers of takedown requests to Google Web search. Similarly, Gray and Suzor (2020) found film and music categories being particularly prone to copyright claims, as well as the Gaming category (which was often a subject to music producers' copyright claims rather than the gaming industry (Matsui, 2016) At the same time, the “sports” category is also heavily policed by YouTube and copyright owners (Gray & Suzor, 2020).

As a result, we argue that in addition to those videos earmarked as “Unavailability due to Copyright Infringement Complaint” it is reasonable to add videos with labels “unknown” or

TABLE 2 Variables used in the mean decrease accuracy plot.

Number of importance in prediction model	Variable
1	Category of video
2	Duration of video
3	Published at
4	View count (number of views a video has)
4	Like count (number of likes the video has)
6	Comment count (number of comments a video has)
7	Dislike count ^a

^aWhile dislike count property was made private by the YouTube API as of December 13, 2021, our variables are derived from the meta-data of YouTube videos from 2019. This means that the property was included in an API response at the moment of collection.

“account deleted” if and only if they have content categories associated with film, music gaming, sports and entertainment. On YouTube, these are the categories: 1—film and animation; 10—music, 17—sports, 20—gaming, and 24—entertainment. This leads to a re-evaluation of takedowns that are most likely copyright-related (cf. Table 3). In conclusion, our best-effort estimate after this multistep assessment is that 2.17% of videos in our sample

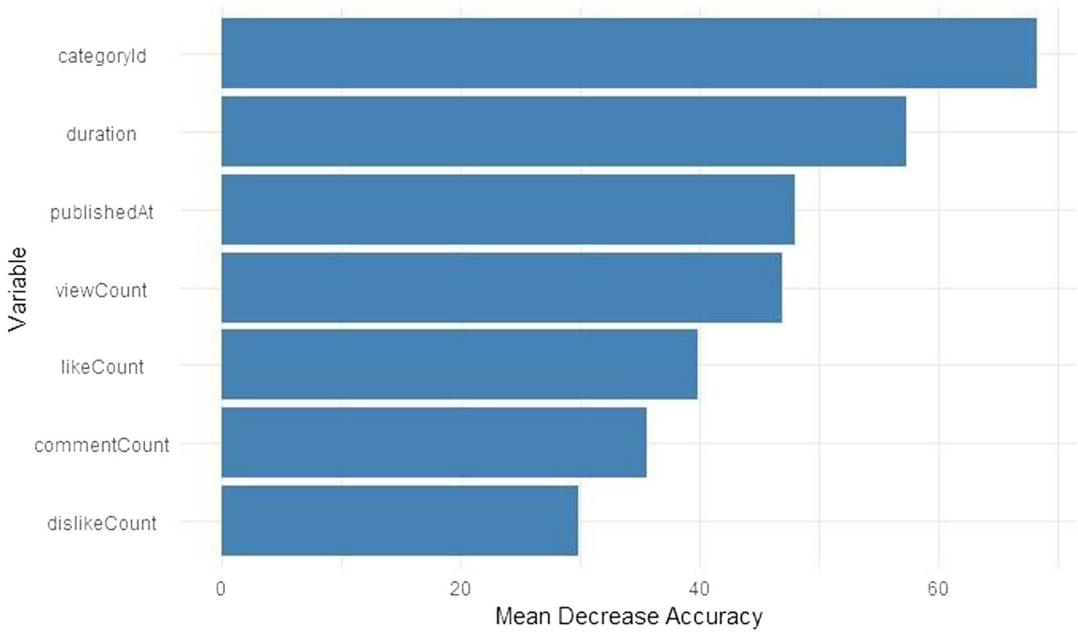


FIGURE 2 The mean decrease accuracy plot model.

TABLE 3 Best-effort estimate on copyright.

	Number of videos	% of videos from sample
Videos in sample	83,676	
Videos unavailable	6311	7.54
Video unavailability due to privacy settings	2774	3.32
Removal due to complaint for false advertising	1	0.00
Deletion of associated YouTube account (categoryID = 1, 10, 17, 20, 24)	175	0.21
Unknown reason (categoryID = all other categories)	1542	1.84
Unavailable without copyright relation	4492	5.37
Unavailability due to copyright infringement complaint	8	0.01
Deletion of associated YouTube account (categoryID = 1, 10, 17, 20, 24)	208	0.25
Unknown reason (categoryID = 1, 10, 24)	1603	1.92
Unavailable with copyright relation	1819	2.17

may have been taken down, both by the platform and by the users themselves, as unavailable due to copyright content moderation.

This result sits right in the middle of existing scholarship on copyright-related takedown rates. While in Gray and Suzor's (2020) study only approximately 1% of all uploaded videos had been removed due to apparent copyright violations, an analysis by Erickson and Kretschmer (2018) of videos highly susceptible to takedowns, such as parodies, revealed with 15.5% a far higher percentage of takedowns that might be copyright-related. The fact that Erickson and Kretschmer have this high estimate might not be surprising: with parody, they focused on content this is highly susceptible to copyright takedown. In addition, they have included all takedowns as potentially copyright-related in contrast to our multistep approach. The difference between the estimate by Gray and Suzor and our own might be related to growing pressure and external regulation of platforms such as the EU CDSMD, which we will investigate in the next section.

Assessing the role of EU regulation in copyright content moderation

Comparing Germany and France on takedowns likely related to copyright

Based on this best-effort assessment of the role and scope of copyright content moderation in takedowns, we have then compared the findings for the two different countries under study (Germany and France) to test for potential early effects of CDSMS implementation on copyright content moderation. For that assessment, we have compared numbers for all content that has been taken down, except the category "Video Unavailability due to Privacy Settings" which mostly stems from security measures rolled out in 2021 which made videos listed as private (YouTube Help, 2021). This yields a baseline of 3535 videos for this cross-country comparison.

The results show remarkable differences between Germany and France (cf. Table 4). In France, there have been more takedowns in general (including videos made private) with 3410 takedowns in comparison with 2901 in Germany. Yet the relative share of copyright-related takedowns is much higher in Germany with almost two-thirds of takedowns (64.19%) being copyright-related in comparison with only a bit over a third (39.62%) in France.

TABLE 4 Difference between takedowns in Germany and France.

Videos unavailable	3535	
Germany	1681	100.00%
Unavailable with copyright relation	1079	64.19%
Unavailable without copyright relation	602	35.81%
France	1855	100.00%
Unavailable with copyright relation	735	39.62%
Unavailable without copyright relation	1120	60.38%
Standard error	0.017	
Z score	14.594	
Significance level (alpha)	0.05	
Confidence level	0.95	

The statistical check (z score test) of these descriptive results confirm that the difference is indeed statistically valid and significant. The calculated z score (14.594) is much larger in absolute terms than the critical value of ± 1.96 , confirming the significance of the difference in proportions. While we cannot prove that this difference is causally linked to CDSMD, yet there is a clear statistical and temporal correlation that answers RQ1: Germany as the country with early CDSMD implementation displays significantly higher shares of copyright-related takedowns than France, which is lagging behind in terms of implementation of the directive.

To contextualize these results, it is important to note that national copyright regimes have always differed between France and Germany, the implementation of CDSMD with regard to timing and substance is not their only difference—but in general, the copyright regimes in France and Germany had been harmonized on a quite high level already before the CDSMD (Sganga et al., 2023). Since Article 17 of the CDSMD, as highlighted by legal scholars (Husovec & Quintais, 2021), is not merely a “clarification” of the existing law, but it changes the law in fundamental ways (Husovec & Quintais, 2021). So, while longstanding differences in the copyright regimes of the two countries might cause different blocking behaviors, we consider the early German implementation of the CDSMD a more plausible explanation for the observed differences.

In addition, cultural attitudes to using copyright infringement works are different in Germany and France. But according to the pre-existing cultural attitudes, we should have seen more infringement attempts, and thus removal, in France and not in Germany. A survey by the European Union Intellectual Property Office (2021) on online copyright infringement in the EU showed that Internet users from France differed from those in Germany in their access to pirated content. While Germany had the second lowest number of accesses to illegally distributed content (including movies, TV, and music) with four accesses per user per month, France was among the countries with much higher regular use of pirated content (European Union Intellectual Property Office, 2021).

There also exist different licensing agreements between platforms and other intermediaries, but in many cases, they are confidential and there are few sources on this. As one example, German case shows that there is already an important licensing agreement between YouTube and German Performing Rights Collection Society and licensing body which had been in place for a while because of differences in remuneration issues (ZEIT ONLINE, 2016). In consequence, as our discussion of alternative explanations runs dry, the early German implementation of the CDSMD yields the most plausible explanation for the observed differences.

Impact on the cultural diversity and YouTube's infrastructure in Germany and France

As a next and final step, we have assessed the impact of YouTube's copyright content moderation in the context of the CDSMD on cultural diversity. For this assessment, we have again worked with the data set of Rieder et al. (2020) as a baseline and compared this with our data new collection in 2022; we have applied the HHI for measuring diversity (see Methods section).

Measure change in diversity for Germany

In Germany, 35 primary categories were used by the owners to identify channels in 2019, while 39 were used in 2022, as indicated in Figure 3.

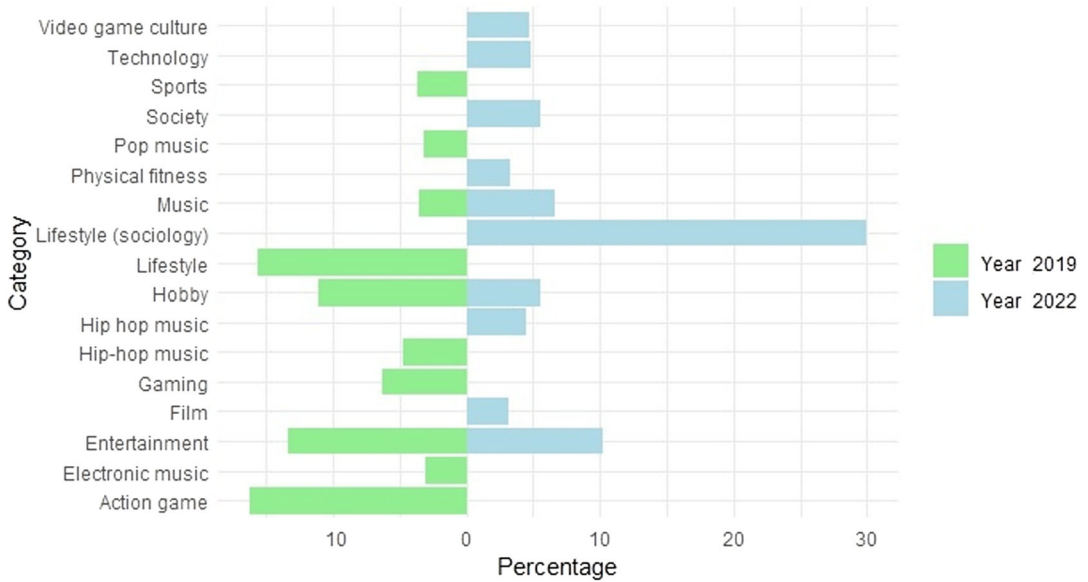


FIGURE 3 Channels distributions by most popular categories in the two samples from 2019 to 2022 in Germany.

In relation to Germany's YouTube channels and using genre as the form of categorization (Moreau & Petlier, 2004), the HHI increased from 958 in 2019 to 1225 in 2022, meaning that cultural production diversity across Germany's YouTube channels fell by 267 points during this period. As a comparison, Benhamou and Petlier (2007) concluded their analysis of the French publishing industry from 1993 to 2003 that the HHI in this industry changed by only 100 points (from 2507 in 1993 to 2409 in 2003). As a result, the observed change of HHI of 267 in our German data set can be considered as relatively large.

Measure change in diversity for Germany

In France, 36 primary categories were used to identify channels in 2019, while 45 were used in 2022, as indicated in Figure 4.

The HHI in relation to YouTube channels in France increased slightly between 2019 and 2022 (from 972 to 1055), meaning that production across the YouTube channels in this country had become less diverse. However, the index changed by only 80 points, far less than the German equivalent (267). Table 5 provides a detailed exposition of the variables and a comparison of the indexes.

Looking at the actual categories, we can observe that the following categories were no longer used by YouTube in 2022: Gaming, Movies, Sports, and TV Shows (cf. Table 6 for a comparison of channel categories and their concentration by year and by country, the full table is displayed in Appendix A). The categories losing a large number of channels between 2019 and 2022 included Action Games, Electronic Music, Entertainment, Hobby, Pop Music, Rock Music, and Role-playing Video Games. Referring to previous studies of copyright takedowns (Gray & Suzor, 2020; Urban et al., 2017), all these categories appear prone to copyright takedowns. Meanwhile, the Lifestyle, Knowledge, and Society categories all gained channels, and some new categories appeared, such as Video Game Culture. The full table in Appendix A also shows that categories such as "Humor," niche music (e.g., "Asian

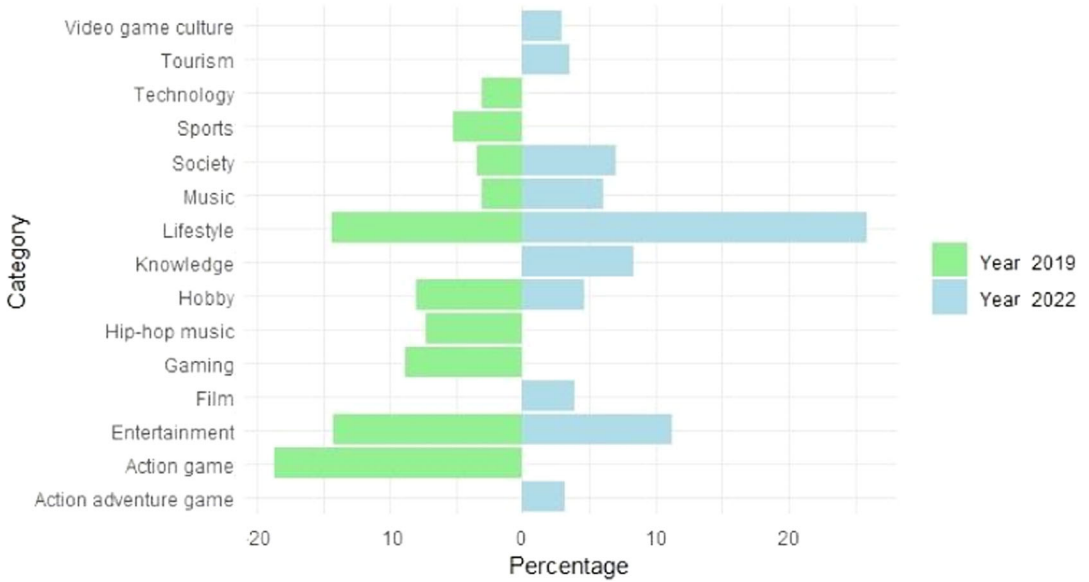


FIGURE 4 Channels distributions by most popular categories in the two samples from 2019 to 2022 in Germany.

TABLE 5 Comparison of supplied HHI on YouTube in Germany and France, 2019 and 2022.

Variety (number of channels categories) and balance (number of channels in each category)							
Supplied	Channels				Supplied HHI		Difference in supplied HHI
	2019		2022		2019	2022	
	Number of channels	Number of categories	Number of channels	Number of categories	HHI	HHI	
Germany	1372	35	1320	39	958	1225	267
France	1382	36	1221	45	972	1055	83

Abbreviation: HHI, Herfindhal–Hirschmann index.

Music”), “Performing arts” and many others barely appeared in the new sample in contrast to being quite present in the 2019 samples.

Thus, in both countries, we observe a decline in the diversity of cultural supply on YouTube, as measured by genre (channel categories), as well as a change in the platform’s infrastructure, possibly related to the categories prone to copyright moderation.

The current research shows how YouTube has changed its infrastructure such that channels now are less diverse than they were in 2019. While this may not be the direct effect of the adoption of Article 17 of the CDSMD, the change in YouTube’s supply in terms of genre may be connected to the prevention of copyright infringement.

Nevertheless, since it is practically impossible to conduct causal research on the platform’s API data, this paper is observational in nature, as are many other studies of content moderation. Without access to the platform’s internal decision-making, it is impossible to know the exact reasons for the change. Such studies always contain confounders—variables that affect both the “treatment” and the “outcome”—so the current

TABLE 6 Changes in YouTube infrastructure of channel categories in Germany and France, 2019–2022 (color coding: red represents categories which were discontinued in 2022; blue represent categories where the number of channels grew in high proportions).

Categories	Germany 2019	Germany 2022	France 2019	France 2022
Action game	224	24	254	25
Electronic music	43	12	30	13
Entertainment	184	135	188	137
Gaming	86	-	117	-
Hobby	153	73	106	56
Knowledge	4	26	6	101
Lifestyle	215	395	190	317
Movies	9	-	30	-
Pop music	44	19	30	20
Rock music	14	5	11	8
Role-playing video game	29	16	32	12
Society	40	74	46	85
Sports	50	-	69	-
TV shows	13	-	32	-
Video game culture	-	62	-	35

authors make no claims regarding causation but do make observations that may lead to further inquiry (Thorburn et al., 2022).

CONCLUSION

Among concerns that both public and private forms of platform governance might reduce cultural diversity, this article has set out to empirically assess effects of external public regulation as well as platforms' infrastructures and practices have an impact on the availability of content and cultural diversity. The context of the CDSMD and its implementation questions with its far-reaching regulations and articulate controversies have offered a pertinent context for investigating such questions. That is why in this study we have sought seek to better understand empirically how CDSMD and its implementation have affected platforms' content moderation, and subsequently the availability and diversity of content.

Despite hard challenges with regard to data access, we are able to identify strong indications of such effects. First, we have developed a more robust estimate for copyright-related takedowns on YouTube by ways of complex data cleaning and calculation. This best-effort estimate results in a take-down rate 2.17% for our sample which is higher share than other general topic studies such as the one by Gray and Suzor (2020). Future studies can build on this multistep methodology to come to more robust estimates for takedown rates than previous studies. Second, a key finding of our research is the significant

difference between Germany and France in terms of content takedowns in categories prone to copyright moderation. Given the otherwise similar country samples, this can be considered an indication of first CDSMD and Article 17 effects, as Germany was one of the first countries to adopt the directive, with France lagging behind many months. Future research can build on these findings and methodologies to assess potential Article 17 effects across a large number of EU. Third, the study found that the content-level supply of cultural diversity, measured by genre (categories) and the distribution of channels over categories (balance), has decreased in both countries, yet with notable differences in size. In Germany, the HHI of the cultural products supply decreased much more than it did in France.

Taken together, these empirical findings provide strong indication of effects of external regulation (here: CDSMD) on platforms' copyright content moderation and, in turn, on the availability of content and cultural diversity. There is a significant amount of copyright takedowns happening; numbers have been rising since 2019, and this effect is significantly stronger in the country that has had early implementation of the directive (Germany); and this has resulted in a decrease of content diversity with again a particularly strong effect in the early CDSMD country.

But there are obvious limitation to our empirical findings. While we can show statistical and temporal correlations, we cannot prove that takedowns and differences are causally linked to the CDSMD and changes in content moderation infrastructures. As Thorburn et al. (2023) note, it remains impossible to conduct causal and in-depth research on platforms' content moderation at the content-level without access to the platform's own extensive data and information.

Another limitation, and the route for future research, might be that the CDSMD theoretically does apply retrospectively to old videos that were uploaded to YouTube before 2021.

However, if a work is posted uploaded to YouTube before 2021 and remains there after the entry to force of the Directive, it could be assumed that whatever changes to the provider's automated rules are made as a result of the Directive will also cover the previously posted work. This could mean anything from monetization to take down rules, and so on. We would suggest that it is very unlikely that YouTube or any other OCSSP would keep two automated systems of copyright infringement in place for works posted before the Directive and another for after. Thus, videos upfolded earlier than 2021, would also be affected. However, this could indeed only be clarified by the platform itself rather than the independent empirical research.

To conclude, this study has shown that with sophisticated methods and processes we can address the questions that we as a society care about: What is the impact of copyright regulation and content moderation on content diversity? But lack of access to social media platforms yields this much more complicated than necessary and prevents specific in-depth studies on the content-level all together. As a result, we need robust mandatory data access clauses in all future platform regulations. The existing initiatives, specifically including Article 40 of the DSA do pave the way, yet implementation questions and practical challenges remain. Finding practical and fair solutions as well as best practices for data access that are not only accessible to researchers at elite and perfectly equipped institutions is a key challenge for policy and research in the next decade, to facilitate the research needed for shaping platforms and their regulation in democratic societies.

ACKNOWLEDGMENTS

This study has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement #870626 as a part of ReCreating Europe project. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

All the new codes for Python used in this research will be available as open access on GitHub. All the data sets are available upon request from the academic institutions.

ORCID

Daria Dergacheva  <http://orcid.org/0000-0001-8773-7934>

Christian Katzenbach  <http://orcid.org/0000-0003-1897-2783>

REFERENCES

- Angelopoulos, C., & European Commission, Directorate-General for Research and Innovation. (2022). *Study on EU copyright and related rights and access to and reuse of scientific publications, including open access: Exceptions and limitations, rights retention strategies and the secondary publication right*. Publications Office of the European Union. <https://doi.org/10.2777/891665>
- Benhamou, F., & Peltier, S. (2007). How should cultural diversity be measured? An application using the French publishing industry. *Journal of Cultural Economics*, 31(2), 85–107. <https://doi.org/10.1007/s10824-007-9037-8>
- Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, 21(11–12), 2589–2606. <https://doi.org/10.1177/1461444819854731>
- Brieske, J., & Peukert, A. (2022). Coming into force, not coming into effect? The impact of the German Implementation of Art. 17 CDSM Directive on Selected Online Platforms (CREATE Working Paper 2022/1). SSRN. <https://doi.org/10.2139/ssrn.4016185>
- Deacon, D., Murdock, G., Pickering, M., & Golding, P. (2021). *Researching communications: A practical guide to methods in media and cultural analysis*. Bloomsbury Publishing. <https://doi.org/10.5040/9781501316951>
- van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society*. Oxford University Press. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- Duguay, S., Burgess, J., & Suzor, N. (2020). Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence*, 26(2), 237–252.
- Erickson, K., & Kretschmer, M. (2018). "This video is unavailable": Analyzing copyright takedown of user-generated content on YouTube (SSRN Scholarly Paper). SSRN. <https://papers.ssrn.com/abstract=3144329>
- European Union Intellectual Property Office. (2021). Online Copyright Infringement in the European Union Music, Films and Tv (2017-2020), trends and drivers (Publication No. TB-01-21-479-EN-N). European Union Intellectual Property Office. <https://doi.org/10.2814/505158>
- Farchy, J., & Ranaivoson, H. (2011). Measuring the diversity of cultural expressions: Applying the stirling model of diversity in culture. UNESCO Institute for Statistics. ISBN 978-92-9189-009-6. <https://www.semanticscholar.org/paper/Measuring-the-Diversity-of-Cultural-Expressions%3A-of-Farchy-Ranaivoson/7514314a6d462408b4be77a44621123c79b9d089>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), eISSN 2053-9517. <https://doi.org/10.1177/2053951719897945>
- Gray, J. E., & Suzor, N. P. (2020). Playing with machines: Using machine learning to understand automated copyright enforcement at scale. *Big Data & Society*, 7(1), 2053951720919963. <https://doi.org/10.1177/2053951720919963>
- Havens, T., & Lotz, A. (2016). *Understanding media industries* (2nd ed.). Oxford University Press.
- Hesmondhalgh, D., Campos Valverde, R., Kaye, D. B. V., & Li, Z. (2023, February 9). *The impact of algorithmically driven recommendation systems on music consumption and production—A literature review [Monograph]*. Leeds. <https://www.gov.uk/government/publications/research-into-the-impact-of-streaming-services-algorithms-on-music-consumption/the-impact-of-algorithmically-driven-recommendation-systems-on-music-consumption-and-production-a-literature-review>
- Husovec, M., & Quintais, J. P. (2021). *How to license Article 17? Exploring the implementation options for the new EU rules on content-sharing platforms*. GRUR International (Issue 4/2021, forthcoming). SSRN. <https://ssrn.com/abstract=3463011>
- Jacques, S., Garstka, K., Hviid, M., & Street, J. (2018). *The impact on cultural diversity of automated anti-piracy systems as Copyright enforcement mechanisms: An empirical study of YouTube's Content ID digital fingerprinting technology* (SSRN Scholarly Paper). SSRN. <https://doi.org/10.2139/ssrn.2902714>
- Katzenbach, C. (2021). "AI will fix this"—The technical, discursive, and political turn to AI in governing communication. *Big Data & Society*, 8(2), 20539517211046182. <https://doi.org/10.1177/20539517211046182>
- Kaye, D. B. V., & Gray, J. E. (2021). Copyright gossip: Exploring copyright opinions, theories, and strategies on YouTube. *Social Media + Society*, 7(3), 2056305121103940. <https://doi.org/10.1177/2056305121103940>

- Keller, P. (2021, December 9). YouTube copyright transparency report: Overblocking is real. *Kluwer Copyright Blog*. <http://copyrightblog.kluweriplaw.com/2021/12/09/youtube-copyright-transparency-report-overblocking-is-real/>
- Lévy-Hartmann, F. (2011). Une mesure de la diversité des marchés du film en salles et en vidéogrammes en France et en Europe. *Culture Méthodes*, 1, 1–16. <https://doi.org/10.3917/culm.111.0001>
- Lin, J., & de Kloet, J. (2019). Platformization of the unlikely creative class: Kuaishou and Chinese digital cultural production. *Social Media + Society*, 5(4), 2056305119883430. <https://doi.org/10.1177/2056305119883430>
- Martin, N. (2019, March 2). *Berliners protest EU's copyright plan*. DW. <https://www.dw.com/en/thousands-in-berlin-protest-eus-online-copyright-plans/a-47753399>
- Martinez-Taboada, F., & Redondo, J. I. (2020). Variable importance plot (mean decrease accuracy and mean decrease Gini) (Version 1). *PLoS One*. <https://doi.org/10.1371/journal.pone.0230799.g002>
- Matsui, S. (2016). Does it have to be a copyright infringement: Live game streaming and copyright. *Texas Intellectual Property Law Journal*, 24, 215.
- McDonald, D. G., & Lin, S.-F. (2004). The effect of new networks on U.S. television diversity. *Journal of Media Economics*, 17(2), 105–121. https://doi.org/10.1207/s15327736me1702_3
- Moreau, F., & Peltier, S. (2004). Cultural diversity in the movie industry: A cross-national study. *Journal of Media Economics*, 17(2), 123–143. https://doi.org/10.1207/s15327736me1702_4
- Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, 20(11), 4275–4292. <https://doi.org/10.1177/1461444818769694>
- O'Brien, D., & Malcolm, J. (2018, June 12). *70+ Internet luminaries ring the alarm on EU copyright filtering proposal*. Electronic Frontier Foundation. <https://www.eff.org/deeplinks/2018/06/seventy-more-internet-luminaries-ringing-alarm-eu-deep-linking-restrictions>
- Pasquale, F. A. (2016). *The Black Box Society: The secret algorithms that control money and information*. University of Maryland Francis King Carey School of Law.
- Phillips, A. A., Walsh, C. R., Grayson, K. A., Penney, C. E., & Husain, F. (2022). Diversifying representations of female scientists on social media: A case study from the women doing science Instagram. *Social Media + Society*, 8(3), 20563051221113068. <https://doi.org/10.1177/20563051221113068>
- Poell, T., Nieborg, D. B., & Duffy, B. E. (2021). *Platforms and cultural production*. Polity.
- Press Release. (2022, May 19). *Copyright: Commission urges Member States to fully transpose EU copyright rules into national law*. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/IP_22_2692
- Quintais, J. P. (2022, April 26). Article 17 survives, but freedom of expression safeguards are key: C-401/19—Poland v Parliament and Council. *Kluwer Copyright Blog*. <https://copyrightblog.kluweriplaw.com/2022/04/26/article-17-survives-but-freedom-of-expression-safeguards-are-key-c-401-19-poland-v-parliament-and-council/>
- Ranaivoson, H. (2010). The determinants of the diversity of cultural expressions. An international quantitative analysis of diversity of production in the recording industry. *Observatorio (OBS*)*, 4(4), obsOBS442010434. <https://doi.org/10.15847/obsOBS442010434>
- Reda F., & Keller P. (2021, June 4). European Commission back-tracks on user rights in Article 17 Guidance. *Kluwer Copyright Blog*. <https://copyrightblog.kluweriplaw.com/2021/06/04/european-commission-back-tracks-on-user-rights-in-article-17-guidance/>
- Rieder, B., Coromina, Ó., & Matamoros-Fernández, A. (2020). *Mapping YouTube*. First Monday. <https://doi.org/10.5210/fm.v25i8.10667>
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press. <https://doi.org/10.2307/j.ctvhrcz0v>
- Sganga, C., Contardi, M., Turan, P., Signoretta, C., Bucaria, G., Mezei, P., & Harkai, I. (2023). *Copyright flexibilities: Mapping and comparative assessment of EU and national sources*. SSRN. <https://ssrn.com/abstract=4325376>
- Siciliano, M. L. (2022). Intermediaries in the age of platformized gatekeeping: The case of YouTube “creators” and MCNs in the U.S. *Poetics*, 97, 101748. <https://doi.org/10.1016/j.poetic.2022.101748>
- Silver, D., Lee, M., & Childress, C. C. (2016). Genre complexes in popular music. *PLoS One*, 11, e0155471. <https://doi.org/10.1371/journal.pone.0155471>
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707–719. <https://doi.org/10.1098/rsif.2007.0213>
- Thorburn, L., Stray, J., & Bengani, P. (2022). *How platform recommenders work. Understanding recommenders*. Medium. <https://medium.com/understanding-recommenders/how-platform-recommenders-work-15e260d9a15a>
- Urban, J. M., Karaganis, J., & Schofield, B. (2017). *Notice and takedown in everyday practice* (SSRN Scholarly Paper). SSRN. <https://doi.org/10.2139/ssrn.2755628>

- Voakes, P. S., Kapfer, J., Kurpius, D., & Chern, D. S.-y. (1996). Diversity in the news: A conceptual and methodological framework. *Journalism & Mass Communication Quarterly*, 73(3), 582–593.
- Witt, A., Suzor, N., & Huggins, A. (2019). The rule of law on Instagram: An evaluation of the moderation of images depicting women's bodies. *University of New South Wales Law Journal*, 42(2), 557–596.
- YouTube Help. (2021). Older unlisted content. Retrieved July 9, 2023, from https://support.google.com/youtube/answer/9230970?visit_id=638244963602839973-2540686715&rd=1
- YouTube Help. (2023). What is a copyright claim. Retrieved July 12, 2023, from <https://support.google.com/youtube/answer/7002106>
- ZEIT ONLINE. (2016, November 1). *YouTube einigt sich mit Gema*. Retrieved September 19, 2023, from https://www.zeit.de/digital/internet/2016-11/urheberrecht-gema-youtube-einigung?utm_referrer=https%3A%2F%2Fwww.google.com%2F

How to cite this article: Dergacheva, D., & Katzenbach, C. (2023). Mandate to overblock? Understanding the impact of the European Union's Article 17 on copyright content moderation on YouTube. *Policy & Internet*, 1–22.
<https://doi.org/10.1002/poi3.379>

APPENDIX A

Table A1

TABLE A1 Changes in YouTube channel categories and HHI in Germany and France, 2019–2022.

	Categories	Germany 2019	Germany 2022	France 2019	France 2022
1	Action-adventure game	1	28	-	39
2	Action game	224	24	254	25
3	Association football	-	1	-	4
4	Basketball	-	-	-	1
5	Boxing	-	-	1	-
6	Business	-	-	-	4
7	Christian music	2	-	7	-
8	Classical music	4	1	-	4
9	Electronic music	43	12	30	13
10	Entertainment	184	135	188	137
11	Fashion	-	-	-	8
12	Film	-	42	-	47
13	Fitness	21	-	2	-
14	Food	12	26	5	24
15	Football	-	-	1	-
16	Gaming	86	-	117	-
17	Golf	-	-	-	1
18	Health	32	6	14	4

TABLE A1 (Continued)

	Categories	Germany 2019	Germany 2022	France 2019	France 2022
19	Hip hop music	65	60	96	20
20	Hobby	153	73	106	56
21	Humor	1	-	-	-
22	Humor	-	1	-	-
23	Independent music	3	2	-	1
24	Jazz	1	-	2	1
25	Knowledge	4	26	6	101
26	Lifestyle	215	395	190	317
27	Motorsport	-	2	-	3
28	Movies	9	-	30	-
29	Music	49	87	42	73
30	Music of Asia	-	1	2	2
31	Music of Latin America	-	-	1	2
32	Music video game	-	-	1	-
34	Performing arts	9	3	8	2
35	Pet	-	2	-	1
36	Physical attractiveness	-	-	-	1
37	Physical fitness	-	44	-	5
38	Politics	-	10	-	12
39	Pop music	44	19	30	20
40	Professional wrestling	-	1	-	-
41	Puzzle video game	2	2	-	1
42	Racing video game	3	-	1	1
43	Reggae	1	-	3	-
44	Religion	-	3	-	-
45	Rhythm and blues	-	3	-	1
46	Rock music	14	5	11	8
47	Role-playing video game	29	16	32	12
48	Simulation video game	1	-	1	1
49	Society	40	74	46	85
50	Soul music	4	1	3	-
51	Sport	-	10	-	30

(Continues)

TABLE A1 (Continued)

	Categories	Germany 2019	Germany 2022	France 2019	France 2022
52	Sports	50	-	69	-
53	Sports game	6	4	6	1
54	Strategy video game	2	12	3	10
55	Technology	42	63	41	33
56	Television program	-	27	-	18
57	Tennis	-	-	-	2
58	Tourism	-	4	-	44
59	TV shows	13	-	32	-
60	Vehicle	-	33	-	11
61	Vehicles	3	-	1	-
62	Video game culture	-	62	-	35
	HHI supply side	0.09585823	0.1225402	0.09725204)	0.1055009
	Total channels	1372	1320	1382	1221
	Total categories	35	39	35	45

Note: $HHI = c_1^2 + c_2^2 \dots c_n^2$

The Herfindahl–Hirschman Index (HHI) assesses concentration within categories of YouTube channels by summing the squared proportions of channels in each category, with higher HHI values indicating greater concentration and lower values suggesting more diversity and competition. It is a measure used to evaluate the distribution of influence or prominence among channels within a specific category.