



# Chilling Tales: Understanding the Impact of Copyright Takedowns on Transformative Content Creators

CASEY FIESLER, University of Colorado Boulder, USA

JOSHUA PAUP, University of Colorado Boulder, USA

CORIAN ZACHER, University of Colorado Boulder, USA

Though current conversations around automated content moderation often focus on newer contexts for this technology such as harassment or misinformation, one other important use case is copyright detection systems such as YouTube's Content ID. This context for content removals online has some unique properties such as its relationship to legal requirements (Section 512 of the Digital Millennium Copyright Act that governs copyright notice-and-takedown), and provides a longstanding example of the intersection between policy, online platform design, and algorithms. Based on qualitative analysis of survey data, this paper examines the impact of DMCA takedowns and automated copyright-related content removals on a community of content creators who primarily share transformative works, remixed content that makes use of copyrighted material. Our data reveals patterns of chilling effects related to prior experiences and perceptions of policies and process. Based on these findings we make recommendations for platform policy and design, education, and advocacy, and discuss the implications of our findings for current and proposed laws.

CCS Concepts: • **Social and professional topics** → **Computing / technology policy**.

Additional Key Words and Phrases: content moderation, copyright, fair use, law, policy, user-generated content, youtube

## ACM Reference Format:

Casey Fiesler, Joshua Paup, and Corian Zacher. 2023. Chilling Tales: Understanding the Impact of Copyright Takedowns on Transformative Content Creators. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 304 (October 2023), 21 pages. <https://doi.org/10.1145/3610095>

## 1 INTRODUCTION

User-generated content platforms engage in a kind of privatized governance over content, expression, and behavior on those platforms, through both policy decisions and technical mechanisms of content moderation. Whether conducted by humans or automation, moderation decisions on platforms are often opaque, and users frequently (perhaps intentionally) have little access to policy processes. Prior work has highlighted issues such as lack of explanation or ambiguity in policies, systematic biases in automation errors, and barriers to or confusion in appeals processes [10, 16, 18, 25, 29, 33, 44, 46]. Though platforms are making governance decisions, there may be little sense of due process as there might be in other contexts—and when the decision is made by automation rather than a person, this difference is often even more stark. As a result, one issue with automated moderation is this concern about due process, and what impact that might have (particularly with respect to false positives and appeals) on the platform's users.

---

Authors' addresses: Casey Fiesler, [casey.fiesler@colorado.edu](mailto:casey.fiesler@colorado.edu), University of Colorado Boulder, Boulder, CO, USA; Joshua Paup, [joshua.paup@colorado.edu](mailto:joshua.paup@colorado.edu), University of Colorado Boulder, Boulder, CO, USA; Corian Zacher, [corian.zacher@colorado.edu](mailto:corian.zacher@colorado.edu), University of Colorado Boulder, Boulder, CO, USA.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2023 Copyright held by the owner/author(s).

2573-0142/2023/10-ART304

<https://doi.org/10.1145/3610095>

Though current conversations around automated moderation often focus on newer contexts for this technology such as harassment, hate speech, or adult content, another domain has a history of quickly demanding that technologies match and classify online content due to strong economic and regulatory interests: copyright [16]. For example, one longstanding content moderation system is YouTube’s Content ID system that automatically flags and removes content posted on the platform that might infringe copyright [16, 39].

Though content moderation in many contexts intersects with broader policy—for example, restrictions on child sexual abuse material, which was part of the impetus for algorithmic removals of adult content on Tumblr [32]—copyright is a particularly strong example since Section 512 of the Digital Millennium Copyright Act requires platforms to remove content accused of infringing. Therefore, automated content removals for copyright infringement are an important example of the intersection of policy, online platform design, and algorithms, and thus the impact that this these intermingled systems have on the users generating content for these platforms.

This paper examines the impact of copyright-related content removals on a community of content creators who traditionally function within a gray area of copyright law (and therefore in a gray area of content moderation). These creators are part of online communities devoted to consumption and creation of fanworks, which are a type of transformative work—that is, new content that makes use of copyrighted material in a transformative way—and make use of many different social media and user-generated content platforms including YouTube [6, 12]. Examples of fanworks include fanfiction (written stories about a character from popular culture, e.g., the continuing adventures of Captain Kirk) or fanvids (remix videos that use media footage edited to music to share an interpretation of the televisual text, e.g., a series of meaningful looks between Captain America and Iron Man to the backdrop of Elvis’ “Can’t Help Falling in Love”). Prior work has examined the complexity of copyright-related decisions in this particular online creative community, particularly around what constitutes fair use of copyrighted content, which is a notoriously gray area. This prior work revealed that misunderstandings of law or platform policy, lack of confidence of their own interpretation, and/or fear of “getting in trouble” sometimes resulted in *chilling effects* where creators decided not to create or share certain kinds of content or use certain platforms [10, 11, 14]. The current study builds off of that work by specifically examining the impact of copyright-related “takedowns”—how they are experienced and responded to—through qualitative analysis of responses to a large-scale survey of fan creators. In doing so, our work also contributes to a growing body of literature examining the impact for creators of content removals or demonetization (copyright-related and otherwise) on user-generated content platforms [1, 18, 22, 25].

In order to understand the impact of these intersecting systems of law, platform policies, and automated moderation in the context of copyright with fan creators as a case study, our research questions included: (1) What types of experiences do fan creators have with copyright online, particularly content takedowns on YouTube? (2) How do fan creators understand and react to takedowns of their work? and (3) What are the impact of these policies and processes on users and their creative work? Thorough our findings we are also able to compare this context to what we know about user experiences with content removals from prior work.

Our findings illustrate how policy confusion, automated moderation errors, and opaque and discouraged appeals processes can chill artistic expression, and reveal patterns of different types of chilling effects. Based on copyright-related content removals as an example, we make policy and design recommendations for both this specific context and for moderation more generally, and also discuss the implications of our findings for both current and proposed public policy.

## 2 BACKGROUND

As internet popularity and accessibility increased throughout the 1990s, so did concerns about the availability of content such as pornography, graphic violence, defamatory statements, and copyright-infringing works, prompting both regulators and online service providers themselves to shape their role in moderating user-generated content. Subsequently, these decisions have come to play a critical role in online community building and engagement, social media user safety, and broader social issues [17, 21]. In addition to explaining the relevant copyright landscape for this study, we will also briefly situate the work within the landscape of content moderation more generally.

### 2.1 Related Work

This study builds on the body of existing research around user experiences with and perceptions of content moderation, particularly with respect to content removals and moderation errors. Studies of specific platforms such as Facebook [44], YouTube [25], Reddit [19], and TikTok [46] show similar patterns of user responses, particularly regarding user perceptions of inappropriate content removals. One survey of users across multiple social media platforms also found that in the absence of authoritative explanation of moderation processes, users developed folk theories of how and why content was removed, frequently expressing frustration and confusion about incomplete knowledge [29]. Both folk theories and studies of specific experiences frequently reveal potential biases (whether automated or human or both) in content moderation systems, including suppression of content from minoritized or marginalized groups [2, 18, 23, 26, 46].

Lack of transparency is also a significant theme across related work. In a study of YouTube's automated content moderation system, Ma and Kou's findings revealed problems with the opacity of algorithmic punishments, which results in significant precarity for creator labor [25]. Creators in turn put in significant effort to attempt to understand and apply practical knowledge of the automated moderation system in order to avoid content removals and demonetization [25]. Creators on both TikTok [46] and Facebook [44] have voiced similar concerns, noting confusion and a lack of transparency around community guidelines violations, which then makes it difficult to avoid subsequent violations. Moreover, on monetized platforms like YouTube and TikTok, the precarity of visibility often has a direct and tangible impact on a creators' income stream, a problem that is exacerbated by opaque algorithmic systems and content moderation decisions [5, 23, 25, 46].

These and related lines of research have also examined user experiences with appeals processes and contesting algorithmic decisions. Many of the participants in Myers West's cross-platform survey reported running into problems during appeals processes, and expressed frustration when there were no resolutions to their problem in processes that they perceived as opaque [29]. Based on their study of content removals on Reddit, Jhaver et al. concluded that a lack of transparency behind content moderation systems can leave users confused and discouraged, and on TikTok, creators noted the confusing and limited options for appeal as well [19, 46]. Across different platforms, users also frequently express frustration at the lack of human interaction during appeals processes [29, 46], which tracks to research that shows that for algorithmic decision-making generally, people often perceive processes that involve human review to be most fair [24]. And overarchingly, Vaccaro et al. identified systematic failures of moderation appeals to be rooted in part in power imbalances, as well as perceived illegitimacy and decisions made in isolation [44].

Prior work has also explicitly examined the impacts of copyright related content moderation. For example, in Kingsley et al.'s broader study of perceptions and impacts of demonetization on YouTube, participants noted the power imbalance in Content ID protecting big corporations more than smaller content creators [23]. Similar to the current study, Brøvig-Hanssen and Jones focused on creators of

a type of remix: music mash-ups [1]. For their survey and interview participants, copyright-related content moderation was an "everyday" experience. These experiences also had a significant impact on remixers' motivation and creative decisions, but less than half had ever explored options to appeal or contest content removals—in part due to fear of more severe consequences [1]. Kaye and Gray also explored YouTube specifically, by analyzing videos on the platform *about* copyright. Their findings revealed overall strongly negative feelings about copyright policies and processes, including imbalance of power, lack of due process, and bad faith claims [22].

The study that this paper describes builds on this prior work, both with respect to understanding more cases of user challenges with and perceptions of content moderation generally, and adding to a small but growing body of work related to copyright, which has not been as deeply covered in the content moderation literature as some other topics such as hate speech and harassment. Kaye and Gray point out the need for more attention to "stakeholders who are often marginalized in copyright policy debates" [22], and this work expands this space, providing support for and nuances on relevant prior work.

## 2.2 DMCA 512 and YouTube Content ID

Following growing regulatory attention to online content in the 1990s, cases such as *Stratton Oakmont, Inc. v. Prodigy Services Co.* (NY Supreme Court, 1995) and *Playboy Enterprises v. Frena* (FL District Court, 1993) signaled that online service providers could be held liable for the user-generated content they hosted on their platforms even if those platforms made good faith attempts to moderate said content. These cases and more would draw the attention of the U.S. Congress to pursue legislative solutions that attempted to strike a balance between freedom of expression and regulation for online service providers. Though many such laws are based in the United States, they typically constrain and impact platform policy and practice globally, in the same way that laws in other countries (e.g., the EU data privacy law GDPR) have far-reaching effects regardless of the country where any given platform operates.

One such early regulatory solution was the Digital Millennium Copyright Act of 1998 ("DMCA") which in part sought to address copyright infringement online. Section 512 of the DMCA ("DMCA 512") outlines liability protections an online service provider may qualify for from copyright holders after hosting infringing content on their platform. The legal provision outlines that upon discovery of copyright infringement taking place on an online platform, the copyright holder must submit a notice to the service provider to initiate takedown procedures. Upon receiving such a notice, the service provider is then required to remove the copyrighted content from their platform, and provide the opportunity for the person who posted the content to appeal. This process of notice and takedown to an online service provider is informally known as a "DMCA takedown." Complying with the notice and takedown of copyright infringement on their platform prevents the online service provider from being held liable for copyright infringement.

Prior to the explosion of social media that necessitated rapid scaling, DMCA 512 may have been more easily enforceable through manual means due to smaller quantities of user-generated content that online platforms were hosting. However, as user-generated content platforms scaled, content moderation in general became a bigger challenge. Some online service providers have turned to automated moderation systems for copyright as well, or other mechanisms for automating takedowns. However, YouTube was one of the first to significantly scale, in part due to legal threats [16]. As a result, YouTube's Content ID is actually *broader* than DMCA 512.

The Content ID system is an automated copyright detection system within the YouTube platform. According to a series of articles on YouTube's Help page,<sup>1</sup> copyright holders can upload references

<sup>1</sup><https://support.google.com/youtube/answer/2797370>

Confirm eligibility      Select type      Review requirements      Provide rationale

Type of fair use  
Educational video

Why is your use of the claimed content an example of fair use? (required)  
Give the claimant as much information as you can about why your video is protected by fair use, fair dealing, or similar exceptions to copyright.  
0/2000

Questions to consider

- How much of the song have you used?
- Have you transformed the song?
- Is your video a substitute for listening to the original song?

This will go to the claimant for review.

Please review the statements below and check the boxes to agree.

☐ My video does not infringe anyone's copyright.

☐ I understand that the claimant will be able to review my video and my dispute rationale.

☐ I understand that filing fraudulent disputes may result in termination of my YouTube account.

Signature (required)

Enter your full legal name (usually a first and last name), not a company or channel name.

Fig. 1. A screenshot of the section of the "Dispute Claim" form for a Content ID claim on YouTube where the creator can explain fair use rationale. The source of this screenshot is a Content ID claim that the first author received on a YouTube video.

of their content to the Content ID system that YouTube can then use to compare against users' videos. These references can take on the form of copyrighted music, movies, television shows, or other forms of video media. If a match is found between a copyright holder's reference and a user's video, YouTube will issue a "Content ID claim" on the video and the copyright holder can choose to either block the video, claim monetization on the video, or track viewership statistics. Content ID also has a monetization option that essentially automates licensing, allowing a copyright owner to claim the revenue from a video rather than having it removed [31]. During multiple levels of the process, the user has opportunities to dispute (see Figure 1) and the copyright holder has opportunities to release the claim or request that the video be taken down entirely. Eventually this process can escalate to the copyright owner issuing a legal takedown notice through YouTube, at which point the user must respond with a counter notice in order to appeal. This process is far from simple. YouTube's Help pages contain a number of flowcharts, though a report for the Electronic Frontier Foundation (EFF) includes new flowcharts representing the lived experiences of users and what steps of the process "actually look like," noting that even copyright experts at NYU Law School had trouble navigating the process [39]. Moreover, too many copyright strikes and removed videos result in the user having their channel and videos removed from the platform.

Perel and Elkin-Koren note the importance of accountability in Content ID's decision-making, considering the large role that YouTube plays in public discourse, but emphasize how challenging it is to understand how it exercises its power due to a veil of proprietary code tied to business interests [31]. However, by all accounts the system is heavily tilted towards the preferences of copyright holders in the same way that DMCA 512 processes are [16, 31, 39]. The EFF report describes the

Content ID system as one that instills a "culture of fear" amongst content creators since challenging a Content ID claim has the potential to result in legal action and financial burdens [39].

Prior work in CSCW has highlighted the challenges that fan creators and other remixers face in online creative communities due to the ambiguity of copyright law and related platform policies, frequently mentioning YouTube specifically, in part because of automated takedowns via Content ID [10, 11, 13]. Moreover, like content moderation research with findings that reveal intense frustration about content removals perceived as unjust [18, 29], fair uses of copyrighted content represent a content moderation "gray area." Our goal in this study was to therefore empirically explore the impacts of copyright-related content removals in the context of automated content moderation by focusing on experiences with the DMCA, Content ID, and YouTube.

### 3 METHODS

#### 3.1 Data Collection

The Organization for Transformative Works (OTW) is a nonprofit organization that was established by fan creators to serve their interests, including maintaining a large, successful fanfiction archive (Archive of Our Own), and engaging in legal advocacy. The first author is a member of the OTW Legal Committee, and in 2017 we partnered with that committee to conduct an online survey about fan experience with and knowledge of copyright. The recruitment was targeted at adults who self-identified as part of transformative fandom (that is, people who consume and/or create fanworks). It is important to note that the stated belief of OTW, and therefore likely by extension the majority of participants in this study, is "that fanworks are transformative and that transformative works are legitimate," with a vision of a future where "all fannish works are recognized as legal."<sup>2</sup> Noncommercial fanworks and other transformative works are largely recognized as fair use in the United States, though it is also the case that fair use is a legal doctrine decided on a case-by-case basis without any blanket exceptions [10]. Additionally, OTW focuses solely on noncommercial fanworks. For the purposes of this study (and supported by our findings), we assume that much of our study population has a good faith belief that fanworks are fair use and therefore do not infringe copyright. In addition, in contrast to similar studies of content creators that touch on monetization [22, 23], the content in question is largely noncommercial and rarely monetized.

OTW disseminated the call for participation via social media channels (e.g., Twitter and Tumblr) and mailing lists. The survey was also approved by the University of Colorado institutional review board, and any participants who did not agree via a consent form to be part of an academic study were not included in our data analysis. A total of 2,376 participants consented and answered some part of the survey; all questions were optional.

The survey contained questions about fandom participation, perceived level of copyright knowledge, copyright-related opinions, and experiences with copyright. The survey also included (as described below) questions specific to YouTube and to content takedowns due to (a) the prevalence of these themes in prior work regarding chilling effects [10][13] and (b) what the OTW legal committee shared about common copyright-related problems for which their help was requested. The survey also asked for demographic information, including age, gender, location, occupation (all open answers), and level of education (multiple choice).

#### 3.2 Data Analysis

Though members of the OTW legal committee conducted analysis of the survey data for reporting back to their community [3], for the current study, the authors of this paper conducted additional

<sup>2</sup>[www.transformativeworks.org](http://www.transformativeworks.org)



analysis of only the responses that related to experiences with content moderation via copyright-related content removals. These responses came from two questions related to the DMCA (which as described above is the typical mechanism for copyright takedowns on user-generated content platforms) and content takedowns specifically on YouTube, and a third question about general copyright experiences where the answers often touched on these topics. In analyzing these responses specifically for the HCI and social computing community, we were able to focus in on issues of content moderation and also *impact* on members of this community. Given the large number of open answer responses, these questions also provided a significant amount of qualitative data. The specific survey questions included:

Q34. Are you familiar with the Digital Millennium Copyright Act (DMCA)?

- No, I have never heard of it.
- I have heard of it, but I do not know what it is.
- Yes, and I think the DMCA is:

Q35. Do you know what to do if your fanwork is the subject of a takedown notice on YouTube?

- I do not know what a takedown notice is.
- I have received a takedown notice and here is what I did do:
- I am familiar with takedown notices but would not know what to do.
- Here is what I would do if I received a takedown notice:

Q44. Have you ever experienced an issue with copyright related to your fanwork? (e.g., your work being taken down, or finding out that someone else copied your work)

- No
- Yes (explain what happened):

Q34 was intended to get a baseline of relevant copyright knowledge from participants. Q35 was intended to provide actual past experiences with copyright-related takedowns as well as hypotheticals for those who had not experienced copyright-related takedowns. We analyzed Q44 as well, because although some respondents reproduced their answers about YouTube experiences, it also provided some relevant non-YouTube DMCA experiences – including those in which the DMCA helped participants to *protect* their work in other contexts.

The total number of responses for the entire survey was 2,376, though since not every participant answered every question, our dataset represents a total of 2005 participants who answered at least one of Q34, Q35, or Q44. In addition to quantitative responses, Q34 included a total of 441 open response answers, Q35 included 462, and Q44 included 278.

In analyzing the data, first we calculated descriptive statistics for multiple choice survey responses; because all questions were optional and not all participants answered every question, we report an N for each of these quantitative findings. Next, we conducted qualitative analysis of open responses to parts of the above questions (totaling 1,181 open responses). To conduct this analysis, we began with two researchers (the first and third authors) independently conducting inductive, open coding on a subset of the data (10%, stratified across questions) and then coming to a consensus on high-level categories to be applied deductively to the remainder of the data. We conducted this analysis and created codebooks separately for each type of response that we analyzed; for example, for respondents who reported experiences with takedowns in Q35, we created and applied codes such as "ignore," "got help," or "counternotice." One researcher conducted the majority of the deductive coding following the creation of these codebooks, and met weekly with a second researcher to discuss, iterate, and make collective decisions on edge cases. During this process we also iteratively updated our codebooks (and coding) when necessary, which allowed us to verify thematic saturation in our codes. Following the application of our codebooks to the entire dataset,

we met to discuss cross-cutting themes across research questions and categories, and then wrote theme memos based on the coded data, which became the primary themes in our findings. Both researchers who conducted this analysis are law school graduates with expertise in both copyright law and fandom.

### 3.3 Participants

The demographics of survey participants (who answered demographic questions) track to other large-scale fandom surveys [12]. We asked participants to self identify their gender, age, location, and level of education. Participants were 84.4% women, 6.8% men, and 5% non-binary or genderfluid, with the remainder providing a range of other responses. They were between the ages of 18 and 72 (mean 28), a majority (59.2%) reside in the United States, and 51.6% have an educational level of at least college. Note that we did not ask participants for their race or sexual orientation, but previous fandom surveys suggest that this population is likely not very racially diverse (majority white) and likely has a large number of LGBTQ+ participants [12]. The vast majority of participants (85.6%) in this sample are fan creators rather than just consumers, many of whom are very active participants in fandom (including writers and artists) who use a range of platforms but are most active on Archive of Our Own and Tumblr. Previous interview studies related to copyright and fan creators have been drawn from similar populations [10][11].

## 4 FINDINGS

Our findings touch on experiences with and perceptions of copyright and copyright-related processes as related to participants' own work. As a reminder, fair use is decided on a case-by-case basis and is often contested—including by copyright owners, who though in actual cases of fair use have no say in the matter, often disagree about whether something is fair use or not. Therefore, though noncommercial fanworks are generally accepted as not infringing [10, 42, 43], even if we had more detail about individual cases in our data we would not be able to know for certain whether a particular participant's content was correctly or incorrectly taken down as a copyright violation. However, because the responses we analyze focus on fanworks in particular, we assumed for the purposes of our analysis that the content being discussed could reasonably be fair use.

Therefore, our findings largely highlight the challenges and problems associated with copyright-related takedowns for content creators with good faith beliefs that their content is not infringing. In contrast to previous work on copyright challenges for remixers and fan creators [10, 13, 14], rather than focusing on the ambiguity of the law itself (e.g., "Is this actually fair use?"), our findings that dive into content removals specifically reveal *process*-related issues, including complications involving the use of automation to make legal judgements and lack of access to due process. First, we briefly detail some descriptive quantitative findings, then qualitative evidence of chilling effects, which we categorize into three major types: policy-based, process-based, and fatalistic. Finally, we end with examples in our data of when takedown process works as intended.

First, recall that because of the connection between Content ID removals and formal DMCA takedowns on YouTube specifically, although a DMCA takedown still has to be triggered manually by a copyright holder, they typically *begin* as an algorithmic decision via Content ID. It is important to note, therefore, that although our survey questions as described above refer to the DMCA and "takedowns," our analysis suggested that many respondents are referring to content removals and/or appeals via Content ID *before* they receive a formal DMCA takedown. Because our findings focus on the impact of copyright-related content takedowns generally, this distinction is not critical to our overall interpretations. However, because of this lack of distinction we also do not make quantitative claims about the prevalence of one or the other in our data. Of the 1,844 total survey respondents who answered Q34 ("Are you familiar with the DMCA?"), only 335 responded that



Question No.	Response	N
Q35(1)	what is a takedown?	249
Q35(2)	have received takedown	110
Q35(3)	would not not know what to do	1168
Q35(4)	would know what to do	402
TOTAL		1903

Table 1. the number of responses for each of the four options for Question 35: Do you know what to do if your fanwork is the subject of a takedown notice on YouTube?

Question No.	Response	N
Q44(1)	no copyright experiences	1605
Q44(2)	have copyright experience	338
Q44(2)	have experience - infringing	100
Q44(2)	have experience - infringed	173
TOTAL		1952

Table 2. the number of responses for Question 44: Have you ever experienced an issue with copyright related to your fanwork?

they were familiar with the law, and out of those, less than a third correctly or partially correctly described it.

Additionally, Q35 gives us an idea of how many participants have actually had experiences with content removals on YouTube specifically, either through formal DMCA takedowns or Content ID claims. First, note that out of the 2,374 total survey respondents, less than 20% reported that they created fanvids, which is the primary type of content that would be shared to YouTube. As Table 1 shows, of the 1903 respondents who answered this question, 110 of them had experienced a takedown; a number of quotes in our findings comes from these 110 respondents.

As Table 2 shows, of the 1952 respondents who answered Q44 about whether they had had a copyright-related experience with a fanwork, 338 respondents answered in the affirmative. 278 of those provided additional details as an open response. As part of our coding process, we also categorized each of those responses for whether the experience was about the respondent being accused of infringing (100), or whether the respondent thought their work had been infringed (173). A small number could not be categorized as either.

The majority of the findings below are based on analysis of the open response answers for Q35(2) and Q44(2), which totals about 400 free responses. Each quote in our findings comes from a unique participant. Quotes are representative of broader themes in our findings, and are edited only for spelling and punctuation.

4.1 Chilling Effects

Regardless of whether they understand the basic practicalities of DMCA takedowns or Content ID, most participants saw these mechanisms mostly as a hindrance to their creativity and legal rights. Our data suggests that this attitude is in part due to the "black box" nature of both the Content ID algorithm and the takedown process itself. Without being prompted to express an opinion, a number of respondents were negative or hostile in explaining the policy – however, in our analysis

we did not perceive this hostility as related to a desire to infringe copyright, but rather rooted in feelings of confusion or injustice. Participants described experiences of fear and confusion about what the state of the law is, cynicism about the law's technical and political legitimacy, and concern that it reinforces the power imbalance between large copyright owners and internet users without the power and expertise to take advantage of systems like Content ID. However, it is important to note that just as we do not know whether any of the takedowns experienced by our participants were correct or incorrect, we also do not know what might have been at the root of any mistakes; our findings are solely based on the perceptions and attitudes of our participants.

The dominant theme across all of the responses we analyzed was that of chilling effects. In the context of the law, a chilling effect describes a situation where speech is suppressed because of a fear of penalty [10][27]. Prior research about fanworks and copyright has confirmed that this attitude generally exists when operating in the legal gray area of fair use [10][13], though our findings point to specific patterns of causes for chilled content, expression, or action on YouTube specifically. Each of these types of chilling effects describes a reason for our respondents to have decided *not* to do something on the platform, whether that is to create or post content at all, or to take action on a copyright claim. *Policy-based chilling effects* are based on YouTube's specific policies or the way they are communicated or instantiated; *process-based chilling effects* are based on uncertainties around or concerns with copyright-related processes such as counternotices; and *fatalistic chilling effects* are based on feelings of helplessness or giving up. These categories of chilling effects help us to understand not only the concrete impacts of law, policy, and automated systems on our participants, but also the constellation of factors that are impacting their creative and copyright-related decisions.

**4.1.1 Policy-based Chilling Effects.** As shown in prior work, fair use is ambiguous enough that many fan creators are concerned about not knowing whether their content is infringing or not even when acting in good faith [10]. However, this concern is amplified on YouTube where their copyright policy makes the process to fight the automated decision seem arduous and even actively discourages doing so.

A number of participants specifically mentioned the policy around takedowns on YouTube in which videos are taken down automatically via Content ID, and those creators are then required to appeal the decision. Some see this policy as favoring the copyright owner or providing no real recourse to the content creator:<sup>3</sup>

*Usually [takedowns] happen because of music as youtube has quite strict views on audio copyrights. There's nothing an author can do about that, the site's policy is that this content is taken down with or without the creator's consent.*

*I knew there was a chance I could contest it, but I thought even if my vids were legal YouTube might not know that and might just tell me they were illegal, so it wasn't worth bothering.*

One participant even explicitly called out YouTube for intentionally creating a policy that went against the interests of creators:

*\*angrily raised my fists at YouTube and swore to take vengeance on their site for selling out in 2005 when my site could have championed the community of 'creators'\*/rant*

Other participants mentioned that a reason they did not engage with the counter-notice process was that they didn't want to take risks or that the policy is "scary":

<sup>3</sup>In this paper, we use the term "copyright owner" to refer to the party who makes a copyright claim, and "content creator" (typically our participants) to refer to the person who received the claim for a piece of (typically remixed) content. However, it is important to note that the content creator also owns the copyright in the new expression of their work.

*I sent back a 'fair use' claim? something like that? the wording is very scary though because they make it sound like they're going to take you down.*

*I just didn't want to risk anything.*

It is also worth noting that the "risks" perceived in our data were not monetary or related to lost income. In contrast to some similar studies [1, 22], monetization was almost never mentioned. Indeed, there are strong social norms in fandom against commercializing fanworks [11]. Instead, fears seemed to be around lost content, account bans, or legal trouble. Similarly, Brøvig-Hanssen and Jones found that mashup producers (also creating content that is often noncommercial) fear that disputing a takedown might expose them to greater risk than if they just accepted it [1].

For DMCA counter-notice specifically, they are filed under penalty of perjury, which could understandably be scary on its own, particularly given that it requires a "good faith belief" that content is not infringing, and fan creators are often unsure about fair use. Upon receiving copyright strikes, YouTube users are required to go through "Copyright School" which at least at once point required watching YouTube's "Copyright School" cartoon.<sup>4</sup> The video (starring a cartoon rabbit) explains the concept of remix, noting that it might require permission from the copyright owner – unless the new work falls under fair use. A voiceover then proceeds to rattle off text from the fair use statute as it appears on the screen and literally crushes the poor rabbit – ending with the suggestion that he "consult a qualified copyright attorney." Then in explaining the counter-notice process, the video highlights that submitting false information will result in the termination of the user's account, and that "if you misuse the process, you could end up in court." A gavel hits the cartoon protagonist in the head as a voiceover says: "You would get in a lot of trouble. That's how the law works."

From making fair use sound like an impenetrable concept that will never be understood, to suggesting that amateur content creators hire lawyers, to emphasizing legal consequences – all of these things appear to be designed to discourage remix. Our data also suggests that our participants are not necessarily aware of the distinction between a DMCA (legal) takedown and a Content ID claim, and as described by EFF's report, the relationship between the last two is less than clear [39]. The dispute-claim form for Content ID also reminds the creator that filing "fraudulent disputes" may result in their account being terminated and requires them to state definitively that their video does not infringe copyright.

In our data, we see that even in cases that could have a very strong fair use argument, creators were afraid to go through a dispute-claim or counter-notice process. And YouTube's "copyright school" is not only seen as potentially insulting, but it further reinforces that "you could get in a lot of trouble" for taking part in this process:

*I think my video was fair use (a small handful of clips from a movie series rearranged to highlight the director's use of racism) but I was too scared to fight it.*

*Currently, my YouTube account bans uploads and has had a banner on it for month about how I need to attend 'copyright school.' I find that just plain condescending and humiliating.*

The imbalance of power between big copyright owners and content creators on platforms like YouTube was a theme across our data, but particularly came out in these discussions of policies seeming "scary" or against the interests of creators. Another type of incident that arose a few times in our data and could be contributing to policy-related chilling effect is the potential for weaponized content moderation due to how third-party flagging works on platform. Some participants had personal experiences with their content being taken down after being flagged

<sup>4</sup><https://www.youtube.com/watch?v=InzDjH1-9Ns>

for copyright infringement, not by the content owner, but by a third party using the process as a weapon for harassment, something that we have seen evidence of in other content moderation contexts [4, 29, 46].

*People actually once reported me on behalf of a person who had copied my work and wanted my stuff taken down? ??? Drama drama drama.*

*Someone didn't like the couple I made a video about and flagged me on YouTube.*

Though these examples are not related to automated moderation, we do know that Content ID and other automated copyright detection systems can be intentionally misused. For example, a video surfaced in 2021 of a police officer playing a Taylor Swift song on his phone in response to a citizen recording him, saying "You can record all you want; I just know it can't be posted to YouTube" [34]. In other words, because YouTube automatically removes content that appears in their Content ID database, he was seemingly confident that he was safe from any recording that would include copyrighted music in the background. Kaye and Gray also heard evidence of notice-and-takedown abuse, describing it as "a central problem with copyright enforcement on YouTube" [22]. Weaponization of copyright to shut down critique is actually one of the reasons that fair use exists (as a safety valve between copyright protection and freedom of expression [10]), but the policy-based chilling effects described here means that the instantiation of this power imbalance can favor chilling expression.

**4.1.2 Process-based Chilling Effects.** Beyond the policies themselves, confusion around and/or the difficulty of the *process* also results in chilling effects. Even when the creators have less anxiety about or are more confident about their rights and the fair use status of their content, some content creators are either unaware that there is a way to fight a takedown, are confused about the process, or simply find it too arduous. Fiesler and Bruckman's 2014 paper about fair use and fanworks include a quote from a fan creator who, after a single takedown, stopped using YouTube altogether and instead kept their fanvids on a password-protected website, because they were afraid of getting into trouble [10]. We saw many similar examples in our data, sometimes because they didn't *know* there was recourse:

*Eventually I took the works down altogether. I was unaware that I might have recourse and so felt this was the only option.*

*I'm not entirely sure what to do still...*

Other participants who received takedown notices were aware of the appeal process – but decided that fighting it would be too much trouble or hassle:

*I removed it. I am sure there's a process to it but I don't have that kind of energy.*

*I complied - the prospect of further hassle is unwanted stress.*

*I just remove the video. Saves LOADS of hell.*

Even participants who were confident that their video was fair use or that were particularly motivated still succumbed to the chilling effect of a seemingly arduous process:

*I felt too unmotivated and busy to do anything but it was definitely unfair because the one taken down was a clip I modified into a parody which is perfectly legal for fair use.*

*I knew it probably wouldn't succeed. And I think I was too lazy. It's a shame though - I spent a lot of time on it and would've liked to keep it up to promote the show since it didn't have enough fans in the West.*

At least one of our participants had their entire account shut down due to confusion with the process after receiving a takedown notice:

*I was young and naive (13? 14?) and did not know how to take the fansubbed anime episodes down. My account was shut down.*

One specific type of potential arduousness of the process relates to privacy concerns. Fandom has very strong privacy norms, including pervasive use of pseudonyms over real names [6]. Though there are a number of reasons these norms exist, including a large LGBTQ+ population with safety concerns and a history of stigma, one origin point is fear over the entire community being vulnerable to copyright law [11]. However, the DMCA counter-notice process requires revealing your real name (on YouTube and elsewhere). As one participant revealed, they were unwilling to go through the process for this reason:

*I've had a couple of pdfs removed by MediaFire, presumably because I used pieces of music. I thought about fighting it, but then didn't feel like exposing my fannish identity too much, so I just used another download service.*

Moreover, even the process of appealing an automated Content ID flag on YouTube requires a signature with a "full legal name," despite this being an internal YouTube policy process and not a legal one (see Figure 1). In cases like these, even if the policy itself is fair, or even if it favored the content creator over the copyright owner, the appeals process itself might instigate chilling effects. This might be in the form of allowing the work (even when confident it was not infringement) to stay down and/or refraining from posting other content on YouTube for fear the same thing would happen. Moreover, this privacy concern could, as prior work has confirmed is already a significant problem, exacerbate content moderation harms disproportionately falling on people from marginalized groups [2, 18, 26].

**4.1.3 Fatalistic Chilling Effects.** Our findings also revealed that even when creators might understand the process and even are willing to engage with it on principle, they choose not to because, essentially, they think there's no point in doing so. This may be because of the power imbalance (e.g., they have lawyers and I don't), because of the fact that they're fighting against an algorithm, because of a general sense of "learned helplessness," or because they know the outcome will always be bad.

Considering that the "Copyright School" video actually suggested that a remixer should talk to a lawyer before posting their remix, it is unsurprising that some participants were anxious about resources that might be required to fight a takedown. For example, one participant, even confident that their work was not infringing, both lacked the necessary resources and thought it was "meaningless" regardless to fight an automated decision:

*I let it stand, because I don't have the time/resources to fight for its posting. This was a short film clip to be used by students in a film class I taught, where the link was private and only available to students, and its intended purpose as educational material was clearly stated; all of which is meaningless because the clip was automatically flagged.*

The concept of learned helplessness applies to situations in which one feels as if a given situation or outcome is unavoidable and that they have no control over changing it; we have seen this reflected in people's attitudes about their data privacy [37]. Similarly, many of our participants expressed a "what's the point?" attitude towards the idea of appealing a decision.

*As much as I wanted to fight it, I figured it'd be a pointless uphill battle and sadly accepted the notice for what it was.*

*I googled stuff but was scared my appeal would fail so I did nothing.*

Some of these comments pointed to a feeling of simply being worn down, that after some number of times being flagged, they simply stopped trying:

*Youtube blocked several videos I'd put together featuring copyrighted music and images. I simply stopped posting anything there. Couldn't find the point and had nothing completely original to post at the time. Also, if you post anything that even resembles something else already on the site, your video will be blocked, even if it's completely original work.*

Another participant used the French term for "a thing that is irreversible" to describe a takedown, noting that they felt this way because they couldn't get any information about why the takedown had occurred:

*Once the takedown was fait accompli, I was notified, but really couldn't get anyone to respond to why.*

In response to Q35's question worded as "here is what I did do" in response to a takedown, another participant just responded with "Cry."

Our findings combined with prior work [1, 13, 22] suggest that the origin of this attitude is a combination of factors—not just personal experiences, but hearing about other people's over a long period of time, as well as some inherent distrust of the intentions of big tech companies like YouTube or an assumption that they will always care more about corporate copyright owners than about their users.

These three types of chilling effects—policy-based, process-based, and fatalistic—are not mutually exclusive, and overall attitudes are likely a combination of these factors. However, the end result is the same — potentially non-infringing content is not being not shared, either from the start or due to removals that are not contested.

The attitudes also share a number of similarities to what we know about user experiences with content removals from prior work, though also reveal complexities related to this context's intersection with the law. For most content moderation decisions, the worst case scenario is a content removal or account ban, but in the case of copyright infringement, there was a cloud of fear of further legal trouble, which exacerbated chilling effects.

## 4.2 Intended Effects

**4.2.1 Successful Appeals.** Though it was a small proportion of our data, we also see examples of when DMCA or Content ID counter-notice work as intended. Typically, these involved the respondent having been very confident about fair use, otherwise having permission, or in at least one case, getting assistance from OTW.

*My video was fair use (a parody), and I filled out their protest form. Then they reinstated my video.*

*On cases where I was infringing (eg, [anime music videos] with copyrighted songs) I just accepted the strike, on one of the cases where I had permission from the band, I emailed YouTube with this and the strike was removed.*

*YouTube sent me a notice regarding music use in a vid. I followed the advice of the OTW in responding, citing fair use practices, and the vid was restored.*

The last example, in which the creator sought out help from the OTW (most likely through an email exchange with the Legal Committee), points to the importance of knowledge, education, and assistance in combating chilling effects.

**4.2.2 Protecting Fanworks.** It is also important to remember that the DMCA as a legal mechanism is not just a tool for "big" copyright owners. Content creators also own rights in the original components of their work. Though the process is not typically as automated with Content ID (where copyright owners are notified when a video is algorithmically matched to their content), anyone can request a DMCA takedown if they believe there is content that infringes their copyright



on a third-party platform. Because we also analyzed open response answers from participants who had "infringed" experiences with copyright (i.e., when they thought someone had infringed on their work), we were able to see a few cases of takedowns working for fan creators as the copyright holders as well, though these were all outside the context of YouTube:

*There have been a few scams in fandom where a for-profit group scrapes AO3 or another archive and puts up the works for its own profit, and sometimes my work was involved in that. I never thought of that as a copyright issue though. As far as I know, those scammy websites were taken down. In one of the cases, I sent a notice myself, but I don't know how it works – I was just following the instructions of the people who did.*

*My fics have been changed into ebooks and distributed where I didn't want them. I submitted DMCA and they were taken down.*

In these cases, the power dynamics were flipped—as was the effort. Contrasting with YouTube, where Content ID makes having work taken down via a simple, automatic process, they had to put in work to allege infringement of their own content. Cases like the ones described above are not uncommon; there have been a number of cases of unauthorized third parties reproducing content from Archive of Our Own without permission of the authors. The OTW Legal Committee also helps fan creators deal with unauthorized copies by providing instructions for issuing DMCA takedown notices.

Though there were so few examples of this in our data, we can only speculate that it is possible that based on privacy issues and concerns about filing counter-notices, these same barriers (e.g., linking a legal name to a fanwork posted under a pseudonym) might also prevent many fan creators from protecting their own copyright by filing DMCA claims themselves. In the past when OTW has issued blanket advice for fan creators on DMCA claims it has mostly been when a large website disseminates a huge number of fanworks without the creators' permission. These cases as well demonstrate the difficulty of being the "little guy" in a copyright fight.

## 5 DISCUSSION AND RECOMMENDATIONS

Our analysis of fan creators' discussion of their own experiences with and perceptions of copyright-related content removals (primarily but not exclusively on YouTube) revealed patterns of chilling effects related to both policy and process. Although the focus in this paper is on copyright, these concerns and experiences mirror some of the problems that arise from content moderation error in other contexts as well. In these cases, chilling effects often result in preemptive self-censorship and/or choosing not to appeal incorrect moderation decisions. Based on our findings, which also strongly support findings from prior work on copyright impacts on content creators on platforms like YouTube [1, 22], we offer a set of recommendations related to platform policy and process, education and advocacy, and copyright policy.

### 5.1 Moderation Policy and Process

Users who interact with other user-generated content on social media platforms frequently express frustration with opacity or confusion in platform rules and moderation processes [33][18][19][15][46]. When a platform is intentionally opaque, it may have good reasons for doing so (e.g. keeping bad actors from "gaming" the system [20]) or it may be a means towards secrecy and obfuscation [33]. Regardless, the result is a frustrating experience for users who are attempting to follow the community rules. As Sarah Roberts put it: "The lack of clarity around platform policies, procedures and the values that inform them lead users to wildly different interpretations of the user experience on the same site, resulting in confusion in no small part by the platforms' own design" [33].

Our findings revealed that *policy-based* chilling effects are the outcome of confusing, overly strict, power-imbalanced, or actively discouraging policies in place around copyright. *Process-based* chilling effects are the outcome of opaque, difficult, or otherwise challenging processes in place for taking action. Our participants described their interactions with copyright-related content removals on YouTube as involving both confusing processes and “scary” policies. Is it of course impossible for us to know whether these descriptions serve as a feature or a bug for the system. For example, one could argue that any platform has a stronger incentive to design a moderation system that favors large copyright owners over individual users for legal or financial purposes. It is entirely possible that the power differential that our findings allude to is in fact intentional, in that it may logically be in YouTube’s best interest if creators do not fight DMCA takedowns or Content ID claims. However, assuming good intentions, there are a number of steps that could be taken to improve processes and to encourage and scaffold appropriate appeals.

As previous work on confusion around online copyright online has pointed out, there is much room for improvement in the clarity, readability, and comprehensiveness of copyright-related policies on user-generated content sites, including how rules and penalties are framed [13]. Our findings suggest that YouTube’s policies specifically contribute to the cloud of confusion and “climate of fear” [43] that often surrounds transformative works. YouTube’s “Copyright School” video frames “how the law works” to a remixer as “you could get into a lot of trouble,” rather than as a form of due process that they could use to assert their rights. Thus, in addition to clear, readable platform policies (which unfortunately is still often not the case [7]), there should also be accurate representations of process for both takedowns and appeals, as well as scaffolding of the appeals process that includes active encouragement for users to appeal if they think their content was inappropriately removed. Moreover, additional barriers, such as the requirement for users to provide their real name(s) when not legally necessary, should be reconsidered.

It is inevitable that there will be mistakes in any automated content moderation system. In the case of copyright claims, remixes will very likely disproportionately yield false positives. Fair use, as a subjective, case-by-case legal doctrine, would be challenging if not impossible to model computationally [10]. As Edward Felten once wrote, “The legal definition of fair use is, by computer scientists’ standards, maddeningly vague” [8]. However, based on current accounts of how Content ID works, it is unlikely to be able to take into account anything but the most simplistic aspects of fair use. Of course, in other contexts such as hate speech or harassment detection, false positives in content moderation may amplify inequities such as racial bias [35]. It is therefore critical that there is a fair, convenient, and transparent appeals process for any moderation-based content removal.

These recommendations based on our findings also track to suggestions made by other researchers who have studied user experiences with content removals, including embracing content moderation gray areas by blurring or tagging content rather than removing it [18], involving users more directly in creating content moderation policy [18], increased transparency and explanation for algorithmic moderation [25], compensation for inappropriately punished creators [25], better designs to scaffold appeals and unburden users [44], and establishing explicit guidelines for content removals to nurture users acting in good faith [19]. Like ours, these recommendations emphasize transparency, compassion, and scaffolding.

## 5.2 Education and Outreach

In addition to chilling effects related to concrete policy and process issues, our findings also revealed more general *fatalistic* chilling effects that result from feelings of helplessness, rooted in assumptions or past experiences—the feeling that, as one of our participants put it, takedowns are *fait accompli*, so why bother? Though changes to policies and processes as described above might

also help with this impact if creators see better outcomes, there are other possibilities for helping creators feel more empowered.

Prior work on copyright in remix and fan creation communities has revealed that, despite a great deal of confusion and even misinformation due to gray areas in the law [13][14], many creators are confident in their understanding of the law and have a better understanding of fair use than the average person [10]. As suggested above, platform processes could better scaffold both the appeals processes and copyright knowledge for users; for example, Wikipedia has a sort of “fair use wizard” for image uploading that assists users with navigating the appropriate copyright information for image uploading to alleviate copyright knowledge burdens [13]. However, there may be minimal incentives for platforms to deploy such structures, so external educational initiatives might also be helpful. In the few examples of successful copyright appeals shared by our participants, they had an understanding of fair use, or in at least one example, sought assistance from OTW.

Kaye and Gray’s study of YouTube videos *about* copyright also found that these videos frequently shared strategies for dealing with copyright issues, including ways to manage risks or even subvert the system [22]. Collective action and community knowledge sharing could also be an important component here, though it also comes with an important caveat: prior work on online copyright discussions has also revealed a significant amount of incorrect information, some of which might even be contributing to chilling effects [13]. Therefore, perhaps partnerships between creators and advocacy groups or experts (such as student law clinics) towards open resources and assistance could help shift the balance of power imbalance between platforms and users. These suggestions hold for content moderation contexts beyond copyright as well; for example, the Online Identity Help Center aims to “help people understand different social media platforms’ policies and rules about content takedowns.”<sup>5</sup>

It is worth noting as well that the OTW also serves as an advocacy organization, frequently submitting amicus briefs for cases that implicate transformative works, and contributing to policy comments and proposals, as well as sending representatives to provide testimony before the U.S. legislature and copyright office [42]. Organizations such as the Electronic Frontier Foundation (EFF) and the Center for Democracy and Technology (CDT) also frequently advocate at the policy level for consumers on topics related to digital rights, including copyright and content moderation [40][38]. Additionally, in terms of addressing policy problems that implicate our field, it has been suggested that HCI researchers could have a larger role in policy advocacy as well, which might include working with law clinics to contribute to policy proposals or with organizations like the Computing Community Consortium to lift a collective voice on policy issues [9].

### 5.3 Public Policy

Although most decisions around content moderation (either policies and/or procedures) are made by individual platforms, these platforms are also constrained by legal requirements. Despite the liability protections of Section 230 in the U.S. and other similar laws, some types of content (in addition to copyright infringement, child sexual abuse material would be another example) cannot be handled solely at the platform’s discretion. As a result, there are some changes that might be made through legislative means, in the U.S. or elsewhere, that could have a significant impact on these policies and processes. Here, we discuss how our findings relate to a subset of relevant current and proposed laws.

**5.3.1 DMCA 512.** With respect to copyright takedowns and DMCA 512 specifically, part of the imbalance of power described in our findings partially derives from how the law is framed and implemented. In 2020, the U.S. Copyright Office released a report documenting the results of a

<sup>5</sup><https://www.oihc.org/>

five year study to evaluate the current effectiveness of DMCA 512; one obvious conclusion of the study was that “no potential solution(s) will please everybody” [30]. However, the problem is that “everybody” as defined by the Copyright Office study seems to include only two categories of stakeholders: online service providers (OSPs, like YouTube) and large-scale copyright owners like movie studios and record companies. Even in concluding that “the notice-and-takedown system as experienced by parties today is unbalanced” (which we agree is true) [30], nowhere do we see the experiences of another relevant category: content creators who may be making fair uses of copyrighted content. The Copyright Office virtually ignored these “transformative creators,” along with other kinds of Internet content consumers and creators in this report.

However, these relevant stakeholders are also deeply impacted by the lack of balance in DMCA 512. The Copyright Office acknowledges that the goal of Section 512 is balancing the needs of all relevant parties, but at the same time, it vastly underestimates the extent that improper takedown notices affect users. The nod to “inaccurate notices” (when non-infringing content receives a takedown request) even focuses on the cost incurred to the OSP—with no acknowledgment of the cost to the non-infringing content creator.

However, as revealed by our findings there is a very real cost to transformative content creators as well, and they should be a relevant stakeholder in policy discussions. Inaccurate takedown notices have a tremendous impact on all creators’ ability to share their work online, and those repeatedly accused of infringing may lose content and accounts as a result. These stakes are particularly heartfelt by small creators, who rely on the reach of platforms to share their content.

Overwhelmingly, the fan creators represented in our survey viewed the lack of balance in the takedown process as stifling their creativity and hindering their ability to share their work. The harm of failure to consider fair use was particularly powerful: Some participants whose works were removed explained that they believed that their works constituted fair use, offered examples of uses that fit squarely within fair use factors, and discussed how removal of those works harmed their individual creativity and ability to participate meaningfully in their community.

An assumption that all users are “infringers” or “pirates” and therefore an adversary to the 512 process does a disservice to the many content creators who reasonably and legally rely on fair use. Instead, they should be considered a stakeholder along with OSPs and copyright holders, such that their benefits and harms are part of the balance as well.

**5.3.2 Future Legislation.** Arguably, a move to create more balance between copyright owners and non-infringing creators as stakeholders (particularly within an automated system) might result in less false positives at the expense of more false negatives. That is, we might see more infringing content on platforms like YouTube. Decisions regarding how to resolve these tensions and calibrate for these types of mistakes is perhaps the grand challenge for content moderation on any platform. It may be that many people would disagree about the relative harm of copyright infringement versus inappropriately chilled expression just as some would disagree about the relative harm of hate speech versus inappropriately chilled expression. Currently, DMCA 512 incentivizes, but does not require, online platforms to adopt automated copyright detection systems. However, a piece of legislation under consideration in the U.S. Congress as of 2022 would change this: the SMART (Strengthening Measures to Advance Rights Technologies) Copyright Act. If passed, this law would make automated tools for identifying copyrighted works a requirement for a huge swathe of the internet. Digital rights advocacy groups warn of “collateral damage” of essentially requiring systems like YouTube’s Content ID [36].

We interpret our findings from this study (as well as related studies [1, 22]) as providing supporting evidence for this concern, and lending further warning against such legally mandated measures. If well-established systems on platforms that voluntarily use automated filters, like YouTube and

Facebook, are repeatedly flagging public domain music and literal static as copyrighted content [40], then it is likely that filters implemented by less well-resourced platforms will have even more errors. In arguing against this proposal, EFF also pointed out that when legally mandated, even the platforms that are currently filtering voluntarily may adjust their mistake calibration to minimize false negatives: “If a company’s risk is only lowered by having a filter, then the company will want a filter that is oversensitive; the danger of a copyright suit brought by a billion-dollar company looms much larger in the risk equation than ruining the livelihood of independent creators” [40]. In other words, adding more legal requirements will shift the balance of power even farther away from creators like our participants and exacerbate many of the problems that our study uncovered.

There is also the potential for new laws that might shift the balance of power in automated copyright takedowns and content moderation in ways that our findings would support. For example, the recently approved Digital Services Act (DSA) in the EU, which is anticipated to be adopted and go into effect by 2024 [45], includes components that may be promising. As part of the law’s transparency requirements, the DSA would force providers to disclose details about their policies, procedures, measures, and tools used for content moderation, including those used for automated decision-making [41]. It would also require platforms to implement mechanisms for affected users to appeal against moderation decisions, in order to mitigate the risks of erroneous blocking of lawful speech [28]. If well implemented, measures like these could help prevent chilling effects. Our findings highlight how important appeal mechanisms are, but also how critical it is that they be not only fair, but also clearly explained and actively encouraged.

It is critical that policymaking includes input from a range of stakeholders. However, as noted above with respect to the recent DMCA proceedings, in the context of copyright, stakeholders beyond big copyright owners and big platforms often do not have a seat at the table. Our findings, as well as those from other recent studies of YouTubers [22] and remixers [1], point to the significant impact that public policy changes could have on content creators—for good or for bad. Our hope is that this and similar lines of research continue, and may help give voice to the “user” in user-generated content.

## 6 CONCLUSION

Content moderation for social media and user-generated content platforms is a significant socio-technical challenge at the intersection of policy, design, values, and algorithms. We illustrated one example of how YouTube and similar copyright detection systems impact potentially non-infringing content creators. Our findings reveal concrete chilling effects on creative expression, and point to ways that we might revise policies and appeals systems to better support good faith content creators in the context of copyright and beyond. Finally, we also considered how our findings might be relevant to proposed legislation that implicates copyright and/or content moderation, and would encourage additional research aimed at aiding policymakers in crafting laws that will consider the needs of users as an important stakeholder group.

## 7 ACKNOWLEDGMENTS

Our thanks to the Organization for Transformative Works for help in fielding this survey, members of the fandom community for participating, and the legal committee (particularly Rebecca Tushnet and Betsy Rosenblatt) for their support and feedback. We also thank reviewers and other members of the research community for their insight into this work.

## REFERENCES

- [1] Ragnhild Brøvig-Hanssen and Ellis Jones. 2021. Remix’s retreat? Content moderation, copyright law and mashup music. *New Media & Society* (2021), 14614448211026059.



- [2] Dasom Choi, Uichin Lee, and Hwajung Hong. 2022. "It's not wrong, but I'm quite disappointed": Toward an Inclusive Algorithmic Experience for Content Creators with Disabilities. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 1–19.
- [3] OTW Legal Committee. 2017. Fan Creation & Copyright Survey: Preliminary Results. *OTW Organization for Transformative Works* (2017). <https://www.transformativeworks.org/fan-creation-copyright-survey-preliminary-results/>
- [4] Kate Crawford and Tarleton Gillespie. 2016. What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18, 3 (2016), 410–428.
- [5] Brooke Erin Duffy, Annika Pinch, Shruti Sannon, and Megan Sawey. 2021. The nested precarities of creative labor on social media. *Social Media+ Society* 7, 2 (2021), 20563051211021368.
- [6] Brianna Dym and Casey Fiesler. 2020. Social norm vulnerability and its consequences for privacy and safety in an online community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24.
- [7] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. 2017. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence*. 18–25.
- [8] Edward W Felten. 2003. A skeptical view of DRM and fair use. *Commun. ACM* 46, 4 (2003), 56–59.
- [9] Casey Fiesler. 2020. Lawful Users: Copyright Circumvention and Legal Constraints on Technology Use. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*.
- [10] Casey Fiesler and Amy S Bruckman. 2014. Remixers' understandings of fair use online. In *Proceedings of the ACM CSCW Conference on Computer Supported Cooperative Work & Social Computing*. 1023–1032.
- [11] Casey Fiesler and Amy S Bruckman. 2019. Creativity, Copyright, and Close-Knit Communities: A Case Study of Social Norm Formation and Enforcement. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–24.
- [12] Casey Fiesler and Brianna Dym. 2020. Moving across lands: Online platform migration in fandom communities. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–25.
- [13] Casey Fiesler, Jessica L Feuston, and Amy S Bruckman. 2015. Understanding copyright law in online creative communities. In *Proceedings of the ACM CSCW Conference on Computer Supported Cooperative Work & Social Computing*. 116–129.
- [14] Katharina Freund. 2016. "Fair use is legal use": Copyright negotiations and strategies in the fan-vidding community. *New Media & Society* 18, 7 (2016), 1347–1363.
- [15] Ysabel Gerrard. 2020. Social media content moderation: Six opportunities for feminist intervention. *Feminist Media Studies* 20, 5 (2020), 748–751.
- [16] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (2020), 2053951719897945.
- [17] James Grimmelman. 2015. The virtues of moderation. *Yale Journal of Law & Technology* 17 (2015), 42.
- [18] Oliver L Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.
- [19] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did you suspect the post would be removed?" Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–33.
- [20] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-machine collaboration for content regulation: The case of reddit automoderator. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 5 (2019), 1–35.
- [21] Jialun Aaron Jiang, Peipei Nie, Jed R Brubaker, and Casey Fiesler. 2023. A trade-off-centered framework of content moderation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 30, 1 (2023), 1–34.
- [22] D Bondy Valdovinos Kaye and Joanne E Gray. 2021. Copyright gossip: Exploring copyright opinions, theories, and strategies on YouTube. *Social Media+ Society* 7, 3 (2021), 20563051211036940.
- [23] Sara Kingsley, Proteeti Sinha, Clara Wang, Motahhare Eslami, and Jason I Hong. 2022. "Give Everybody [...] a Little Bit More Equity": Content Creator Perspectives and Responses to the Algorithmic Demonetization of Content Associated with Disadvantaged Groups. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–37.
- [24] Henrietta Lyons, Senuri Wijenayake, Tim Miller, and Eduardo Velloso. 2022. What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 1–15.
- [25] Renkai Ma and Yubo Kou. 2021. "How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
- [26] Brandeis Marshall. 2021. Algorithmic misogynoir in content moderation practice. *Heinrich-Böll-Stiftung European Union* (2021).



- [27] Emily Meyers. 2006. Art on Ice: The Chilling Effect of Copyright on Artistic Expression. *Columbia Journal of Law & the Arts* 30 (2006), 219.
- [28] Jorge Morais Carvalho, Francisco Arga e Lima, and Martim Farinha. 2021. Introduction to the Digital Services Act, Content Moderation and Consumer Protection. *Revista de Direito e Tecnologia* 3, 1 (2021), 71–104.
- [29] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383.
- [30] United States Copyright Office. 2020. Section 512 of Title 17: A Report of the Register of Copyrights. <https://www.copyright.gov/policy/section512/section-512-full-report.pdf>
- [31] Maayan Perel and Niva Elkin-Koren. 2015. Accountability in algorithmic copyright enforcement. *Stanford Technology Law Review* 19 (2015), 473.
- [32] Elena Pilipets and Susanna Paasonen. 2022. Nipples, memes, and algorithmic failure: NSFW critique of Tumblr censorship. *New Media & Society* 24, 6 (2022), 1459–1480.
- [33] Sarah T Roberts. 2018. Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday* 23, 3 (2018).
- [34] Mathew Rockford. 2021. US officer plays Taylor Swift song to try to block video. *BBC News* (Jul 2021). <https://www.bbc.com/news/technology-57698858>
- [35] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 1668–1678.
- [36] Riddhi Setty. 2022. Online Copyright Piracy Debate Ramps Up Over Proposed Legal Fix. *Bloomberg Law* (Mar 2022). <https://news.bloomberglaw.com/ip-law/online-copyright-piracy-debate-ramps-up-over-proposed-legal-fix>
- [37] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 2347–2356.
- [38] Dhanaraj Thakur and Emma Llansó. 2021. Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis. *Center for Democracy and Technology* (May 2021). <https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/>
- [39] Katharine Trendacosta. 2020. Unfiltered: How YouTube's Content ID Discourages Fair Use and Dictates What We See Online. *Electronic Frontier Foundation* (Dec 2020). <https://www.eff.org/wp/unfiltered-how-youtubes-content-id-discourages-fair-use-and-dictates-what-we-see-online>
- [40] Katharine Trendacosta. 2022. The Latest Threat to Independent Online Creators Is the Filter Mandate Bill. *Electronic Frontier Foundation* (Apr 2022). <https://www.eff.org/deeplinks/2022/04/latest-threat-independent-online-creators-filter-mandate-bill>
- [41] Aina Turillazzi, Mariarosaria Taddeo, Luciano Floridi, and Federico Casolari. 2023. The digital services act: an analysis of its ethical, legal, and social implications. *Law, Innovation and Technology* 15, 1 (2023), 83–106.
- [42] Rebecca Tushnet. 2011. Remix nation. *Commun. ACM* 54, 9 (2011), 22–24.
- [43] Rebecca Tushnet. 2014. 'I'm a lawyer, not an ethnographer, Jim': Textual Poachers and Fair Use. *The Journal of Fandom Studies* 2, 1 (2014), 21–30.
- [44] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What It Wants" How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.
- [45] Yasmina Yakimova. 2022. Digital services: Landmark rules adopted for a safer, Open online environment: News: European parliament. <https://www.europarl.europa.eu/news/en/press-room/20220701IPR34364/digital-services-landmark-rules-adopted-for-a-safer-open-online-environment>
- [46] Jing Zeng and D Bondy Valdovinos Kaye. 2022. From content moderation to visibility moderation: A case study of platform governance on TikTok. *Policy & Internet* 14, 1 (2022), 79–95.

Received January 2023; revised April 2023; accepted July 2023