**Testimony of Meredith Rose**
**Senior Policy Counsel**
**Public Knowledge**

**Before the**
**House Judiciary Committee**

# "Copyright and the Internet in 2020: Reactions to the Copyright Office's Report on the Efficacy of 17 U.S.C. § 512 After Two Decades."

**September 30, 2020**

Chairman Nadler, Ranking Member Jordan, members of the committee, thank you for inviting me to testify today on this always-important topic.

Two hundred and twenty-nine million Americans use the internet each day.[1] That's 229 million American adults using the internet to work, worship, connect with family and friends, receive healthcare, consume and discuss the news, and organize political action each and every day. The laws we debate here set the rules for that speech. The ability of these 229 million users to speak freely online must be the first motivating priority of any reform to copyright liability. While we commend the Copyright Office's herculean effort to comprehensively evaluate Section 512, we were alarmed to see the resulting Report dismiss the concerns of everyday users. The Office's analysis performed a familiar sleight-of-hand by presenting user interests as co-extensive with those of platforms, effectively erasing free speech concerns from its analysis.

Congress must abandon the idea that copyright debates are mere sniping between rightsholders and platforms. The speech interest of every American internet user is directly in the crossfire. If we are to strike any sort of "new balance," it must center our nation's 229 million internet users and their ability to speak freely -- not merely the administrative convenience of major industries.

[1] U.S. Census Bureau, *QuickFacts: United States*, (last visited June 1, 2020), https://www.census.gov/quickfacts/fact/table/US/PST045219 (The total U.S. population is estimated at 328,239,523; 77.6%, or 254,713,870 are over 18.); Monica Anderson, Andrew Perrin, Jingjing Jiang & Madhumitha Kumar, *10% of Americans Don't Use the Internet. Who Are They?*, Pew Res. Ctr., (April 22, 2019) https://www.pewresearch.org/fact-tank/2019/04/22/some-americans-dont-use-the-internet-who-are-they/ (If 10% of U.S. citizens over 18 do not use the internet, then 90% or 229,242,483 do).

# I. The risks to users' speech online

## A. ISP disconnections

Alarmingly, stakeholders such as RIAA, MPA and AAP[2] contend that Section 512 grants them the right to demand that an entire household's internet access be terminated, based purely on accusations of copyright infringement. It goes without saying that Congress should not be making it easier for private third parties to unilaterally terminate a household's ability to participate in modern society.

The Federal Communications Commission has found that Americans use broadband "for every facet of daily life."[3] The current pandemic has driven Congress to emphasize the role of broadband in Americans' work, education, social lives, and health care services.[4] Broadband providers have pledged not to cut off people's broadband for non-payment, and Congress has proposed several bills designed to not only protect broadband as an essential communications service, but also to expand access and affordability. Despite this, it is the position of large rightsholders that their unvetted allegations of a civil offense are sufficient to cut an entire household off from the internet. Alarmingly, courts have largely gone along with this argument. This provision, before all others, is sorely in need of revision.

It is hard to overstate the outsize role that broadband access has adopted in the 22 years between the DMCA's passage and today. Americans use broadband to work remotely, attend classes, access critical medical care, consume essential news and information, and socialize. First

[2] U.S. Copyright Office, *Section 512 of Title 17: A Report of the Register of Copyrights* at 98, fn 520 (May 2020) [hereinafter USCO § 512 Report], https://www.copyright.gov/policy/section512/section-512-full-report.pdf.
[3] *Inquiry Concerning the Deployment of Advanced Telecommunications Capability to All Americans in a Reasonable and Timely Fashion, and Possible Steps to Accelerate Such Deployment Pursuant to Section 706 of the Telecommunications Act of 1996, as Amended by the Broadband Data Improvement Act*, GN Docket No. 14-126, 2015 Broadband Progress Report and Notice of Inquiry on Immediate Action to Accelerate Deployment, 30 FCC Rcd. 1375, 1377 ¶ 2 (2015).
[4] *See COVID-19 Broadband Bills*, Public Knowledge (Current as of June 1, 2020) https://www.publicknowledge.org/covid-19-broadband-bills/.

responders use broadband to communicate life-saving information to local residents, and small- to medium-size businesses use broadband to access global markets that are critical to staying afloat.[5] This importance cannot be reconciled with the broad interpretation of Section 512(i) as requiring that ISPs adopt policies that provide for the termination of subscribers upon repeat accusations of infringement.[6]

Much of this disconnect is due to outdated statutory terminology. Put simply, "internet service providers" meant something very different in 1998 than it does in 2020. Modern internet service providers have two distinct functions: the interactive software-level component that connects and routes traffic to the broader internet, and the physical infrastructure over which the traffic flows. In 1998, "internet service providers" were strictly software-layer services such as America Online and CompuServe, which operated over infrastructure provided by the existing telephone network. In 1998, termination from an "internet service provider" meant that a customer had to uninstall American Online and subscribe to any of its software-layer competitors. In short, when the DMCA was written, ISPs were edge services that operated in a competitive market and operated over a separately-owned, regulated common carrier. Congress did not suggest that the operator of the infrastructural component (i.e. the legacy telephone network) could be held liable for copyright infringement.

In 2020, however, the software and infrastructure have come under the same roof; modern ISPs both route traffic *and* own the cable (or fiber) over which the traffic flows. This radically alters the stakes of "subscriber termination," as terminating an account bars the subscriber from the *physical network.* This is particularly dire in light of the current ISP market,

---

[5] Robert Pepper et al., *Cross-Border Data Flows, Digital Innovation, and Economic Growth*, The Global Info. Tech. Rep. 40, 41 (2016), http://www3.weforum.org/docs/GITR2016/WEF_GITR_Chapter1.2_2016.pdf; FCC, *Connecting America: The National Broadband Plan* 313 (March 17, 2010), https://transition.fcc.gov/national-broadband-plan/national-broadband-plan.pdf

[6] *BMG Rights Mgmt. v. Cox Communs.*, 881 F.3d 293, 302 (4th Cir., 2018).

which severely lacks competition. More than 100 million Americans live in homes serviced by only *one* broadband provider.[7] Forty-two million lack access to *any* wired or fixed wireless broadband, instead relying on limited mobile or satellite connectivity.[8] Only 27% of census blocks have access to more than two broadband providers at standard (25 Mbps) speeds; fewer than 2% have competitive access to 100 Mbps speeds.[9]

While deployment and speeds have improved marginally over time, competition has not. The reasons for this are numerous and well-documented: truly high-speed wired broadband is only feasible over fiber and coaxial cable; the DSL providers who once provided a level of competition to cable are increasingly irrelevant; and mobile broadband remains a complement, not a substitute, to wired household broadband for the vast majority of users. Thus, for most households, being cut off from wired broadband means losing the kind of internet access necessary for those public policy reasons -- school, work, and healthcare -- that Congress and the FCC have consistently advanced its adoption.

It is also questionable (both as a legal and policy matter) whether the act of providing broadband access should *ever* give rise to any form of secondary liability from which a provider must be shielded. The law does not specify how ISPs are supposed to obtain knowledge of repeat infringers; they are not required to accept DMCA takedown notices, as other online service providers are, since they do not actually host any material. ISPs also do not (and should not) have a general duty to monitor and track their users' activity. Private allegations of civil offense

---

[7] Christopher Mitchell, *Repealing Net Neutrality Puts 177 Million Americans at Risk*, Community Networks (December 11, 2017) https://muninetworks.org/content/177-million-americans-harmed-net-neutrality.

[8] John Busby, Julia Tanberk et al, *FCC Reports Broadband Unavailable to 21.3 Million Americans, BroadbandNow Study Indicates 42 Million Do Not Have Access,* BroadbandNow (February 3, 2020), https://broadbandnow.com/research/fcc-underestimates-unserved-by-50-percent.

[9] FCC, *Internet Access Services: Status as of December 31, 2016,* fig. 4 (February 2018) https://transition.fcc.gov/Daily_Releases/Daily_Business/2018/db0207/DOC-349074A1.pdf. The FCC's more recent report contains less reliable figures as it includes satellite broadband, which is not an adequate substitute for terrestrial fixed connections, in the higher speed tiers. *See* Internet Access Services: Status as of December 31, 2017, fig. 4 (August 2019), https://docs.fcc.gov/public/attachments/DOC-359342A1.pdf .

have never been -- and should never be -- sufficient reason to compel utility companies to cut off a customer's water or power. Within the context of copyright liability, modern-day ISPs are more akin to the power company that powers a user's computer than they are to a file-sharing service or streaming site.

### B. Bad notices

DMCA takedown notices are extraordinarily powerful tools with a documented history of abuse. A DMCA takedown has unparalleled power in the online ecosystem -- the power to unilaterally, and nearly instantaneously, erase speech from the internet. Even when challenged, the law requires that the speech stay down for up to 14 days.

This power would be concerning even in a well-governed system. However, by any measure, the DMCA notice-and-takedown scheme is *not* well-governed. It suffers from a disproportionate number of bad notices that hide non-infringing speech and information from public view, affecting the ability of users and creators to use the internet for free expression and creativity. A quantitative study of more than 108 million takedown requests revealed that approximately 30% of requests were "potentially problematic," and further, that 4.5 million of the problematic requests were "fundamentally flawed."[10] Common causes include anticompetitive behavior,[11] a misunderstanding or misapplication of copyright,[12] and lack of useful identifying information for either the infringed-upon work or the allegedly infringing material.[13] The *kinds* of bad notices vary by platform, and reflect the particulars of its user base;

---

[10] Jennifer M. Urban, Joe Karaganis, & Brianna Schofield, *Notice and Takedown in Everyday Practice* 2 (March 22, 2017) (UC Berkeley Pub. L. Res. Paper No. 2755628 ) [hereinafter Urban Report], https://ssrn.com/abstract=2755628.

[11] *See* U.S. Copyright Office, *Section 512 Study: 9th Circuit Public Roundtable* 248 (May 13, 2016) (Testimony of Stephen Worth, Assoc. General Counsel of Amazon.com), https://www.copyright.gov/policy/section512/public-roundtable/transcript_05-13-2016.pdf.

[12] *See, e.g.,* Google, Additional Comments Submitted in Response to U.S. Copyright Office's Nov. 8, 2016, Notice of Inquiry 9–10 (February 21, 2017), https://www.regulations.gov/document?D=COLC-2015-0013-92487 ("[W]e explain at the appropriate step in our form that merely being the subject of a photo does not give one a copyright interest in the photo. In our experience, this warning dramatically cut down on the number of misguided notices.").

on Amazon's Kindle Direct, for example, approximately half of DMCA takedown requests are not infringement-driven, but are instead attempts by authors to remove competitors' books from the rankings.[14] Bad notices are, by any measure, pervasive, and have a substantial aggregate impact on user speech.

### 1. Sources of bad notices

Bad notices stem from a variety of sources that range from technical errors to deliberate bad faith. Generally, they can be broken down into four categories: misuse of copyright, abuse of the DMCA takedown procedure for non-copyright ends, technical flaws, and algorithmic defects. Similarly, the goals and motivations behind bad notices can range from political censorship, to innocent error, to overzealous enforcement. Even at their most granular, each category of bad notice still accounts for millions of problematic takedowns.[15] Any solution to address the bad notice problem requires an understanding of these categories and how they occur.

Copyright misuse occurs when a notice sender leverages the notice-and-takedown process to remove content that incorporates or references their work, but is obviously noninfringing or fair use. For example, in 2019, several unreleased seasons of Starz shows and three episodes of American Gods were leaked to the public via a Russian streaming site. Starz used the DMCA takedown process to remove tweets and articles that reported on the leak, even though the coverage did not itself contain any infringing material.[16] By doing so, Starz was able to leverage the DMCA process to censor legitimate -- if embarrassing -- journalism. This type of

[13] Urban Report at 90 (noting that, by conservative estimates, 4.6% of notices contain incorrect or missing information about the allegedly infringed work or allegedly infringing material).
[14] *See* Testimony of Stephen Worth, Assoc. General Counsel of Amazon.com, *supra* note 11. ("[W]ith Kindle Direct publishing, authors routinely try to climb to the top spot in their category . . . by issuing bogus notices against higher ranking titles. And this for us actually accounts for more than half of the takedown notices that we receive.").
[15] Urban Report at 96.
[16] Ernesto, *Starz Goes on Twitter Meta-Censorship Spree to Cover Up TV-Show Leaks (Updated)* TorrentFreak (April 15, 2019) https://torrentfreak.com/starz-goes-on-twitter-meta-censorship-spree-to-cover-up-tv-show-leaks-190415/.

takedown also catalyzed copyright law's most recent scandal, in which two authors both pulled unprotectable stock elements from a popular fanfiction trope.[17] In response to a new competitor in the niche genre, one author had her publisher issue takedowns against her competitor's work across several online retail sites, claiming an infringed-upon interest in stock elements which the author later admitted she had not created.[18]

A similar strain of abuse occurs when a claimant issues takedowns to remove or temporarily disable unfavorable content for reasons wholly unrelated to copyright. One of the most notorious forms of this is a practice known as "backdating." In order to remove or hide content, the actor will make a copy of the content and post it on an obscure site, backdating the copied material to a time before the original post. They will then issue takedowns against search engines and other indexes, forcing removal of the unfavorable original from search results, while ensuring that the fraudulently backdated copy remains far enough down the results to be functionally obscured. News outlet Benzinga was a victim of this exact practice after it published an article about the financial difficulties faced by Amira Nature Foods, a publicly traded company.[19] Other groups, including the Church of Scientology, have used groundless takedown claims to censor criticism and harass former members.[20] Repressive regimes across the world, from Russia[21] to Ecuador,[22] have become adept abusers of the DMCA's notice-and-takedown regime to stifle critics and suppress coverage of human rights violations.

---

[17] Alexandra Alter, *A Feud in Wolf-Kink Erotica Raises a Deep Legal Question*, N.Y. Times (May 23, 2020) https://www.nytimes.com/2020/05/23/business/omegaverse-erotica-copyright.html.

[18] *See, e.g., Atari, Inc. v. N. Am. Phillips Consumer Elecs. Corp.*, 672 F.2d 607, 616 (7th Cir. 1982); *Walker v. Time Life Films, Inc.*, 784 F.2d 44, 50 (2d Cir. 1986) ("Elements such as drunks, prostitutes, vermin and derelict cars would appear in any realistic work about the work of policemen in the South Bronx.").

[19] Andrea Fuller, Kirsten Grind & Joe Palazzolo, *Google Hides News, Tricked by Fake Claims*, Wall St. J. (May 15, 2020) https://www.wsj.com/articles/google-dmca-copyright-claims-takedown-online-reputation-11589557001.

[20] Eva Galperin, *Massive Takedown of Anti-Scientology Videos on YouTube*, Electronic Frontier Found. (September 5, 2008), https://www.eff.org/deeplinks/2008/09/massive-takedown-anti-scientology-videos-youtube.

[21] Fuller, *supra* note 19.

[22] Alexandra Ellerbeck, *How U.S. Copyright Law Is Being Used to Take Down Corr*ea's *Critics in Ecuador*, Comm. to Protect Journalists (January 21, 2016), https://cpj.org/2016/01/how-us-copyright-law-is-being-used-to-take-down-co/.

Most bad notices are the result of technical errors which, despite being technical in origin, nevertheless undermine the fundamental due process protections built into Section 512. Two of the most substantively important requirements—that a takedown notice contains sufficient information about the allegedly infringed work ("AIW")[23] and allegedly infringing material ("AIM")[24]—are often unmet.[25] The same study that found problems with 30% of all takedown notices also discovered that it was difficult to identify the AIM in 13.3% of requests, and difficult to identify the AIW in 6% of requests.[26] Moreover, notices covering multiple claims do not always include clear details on the location of the allegedly infringing works. This has resulted in substantial, costly litigation over whether rights-holders or OSPs bear the cost of identifying infringing work.[27]

Finally, many bad notices can be pinned squarely on the rise of algorithmic monitoring and enforcement. Though the limitations of algorithms are discussed more extensively below, some examples may be illustrative. In one case, NBC issued automated takedowns against NASA's SpaceX launch livestream--because NBC was using the same feed on its own network, under a license (ironically) from NASA.[28] In another, algorithmic enforcement "blocked a 10-year-old boy's self-authored original video starring his LEGO mini-figures and garbage truck despite the fact that he used royalty-free music."[29] Ultimately, user speech and online ecosystems cannot sustain a system that defaults uniformly in favor of those issuing takedown notices.

---

[23] 17 U.S.C. § 512(c)(3)(A)(ii).
[24] § 512(c)(3)(A)(iii).
[25] Urban Report at 93.
[26] *Id.* at 94.
[27] *Id.* at 93; *see*, *e.g.*, Perfect 10, Inc. v. Google, Inc., No. CV 04-9484 AHM SHX, 2010 WL 9479060 (C.D. Cal. July 30, 2010), aff'd, 653 F.3d 976 (9th Cir. 2011).
[28] Chris B - NSF (@NASASpaceflight), Twitter (May 28, 2020, 9:46 AM), https://twitter.com/NASASpaceflight/status/1266002935051403264?s=20
[29] Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 Stan. Tech. L. Rev. 473, 476 (2016), https://law.stanford.edu/wp-content/uploads/2016/10/Accountability-in-Algorithmic-Copyright-Enforcement.pdf.

### 2. Lack of redress

There are no disincentives, either in the statute or the common law, against filing malicious notices. Section 512(f), which Congress included to deter abuse *ex ante* by providing penalties for bad notices, has been rendered dead letter -- an outcome which, it is worth noting, is endorsed with some enthusiasm in the Copyright Office's 512 Report.[30] Users whose speech has been improperly removed lack any meaningful redress *ex post* as well. Counter-notices are subject to a waiting period of up to 14 days, a duration that can be lethal to time-sensitive speech including news reporting, documentation of human rights abuses, political speech, public debate, and critique. For individuals who make their living through their online speech, that fourteen days represents the loss of both direct income and relevance. Moreover, the mere act of filing a counter-notice opens the user up to a potentially frivolous lawsuit. It is perhaps no surprise that platforms have reported counter-notice rates between 4.7% and 0.02%.[31]

### 3. Platform responses to bad notices

In the absence of meaningful statutory safeguards, some platforms have sought to screen out abusive or defective notices, including by requesting missing information, clarification of ambiguous notices, or additional information about the nature of the claim prior to processing. As noted above, different communities, and the platforms on which they congregate, have different use cases for copyrighted content, and thus different risk profiles for use (and misuse) of DMCA notices. For example, the Hugo Award-winning Archive of Our Own, maintained by the nonprofit Organization for Transformative Works, hosts more than four million works which

[30] The Copyright Office believes that the only case which provides even a modest nod toward § 512(f)'s enforceability—*Lenz v. Universal Music Group Corp.*, 815 F.3d 1145 (9th Cir. 2016)—was wrongly decided for placing potential liability on rightsholders. *See* USCO § 512 Report at 5.

[31] *See* Senate Committee on the Judiciary Subcommittee on Intellectual Property, *Is the DMCA's Notice-and-Takedown System Working in the 21st Century?* (June 2, 2020) (Testimony of Abigail A. Rives, Intellectual Property Counsel, Engine Advocacy and Research Foundation) https://www.judiciary.senate.gov/imo/media/doc/Rives%20Testimony.pdf.

remix major media properties and one another.[32] Other sites, such as TikTok, base their core functionality around users' ability to share, remix, and build upon one another's work, attracting users specifically because of that function. And some sites, such as ecommerce platforms, are more at risk for abusive or anticompetitive takedown notices that could substantially prejudice the economic interests of merchants or artists using the platform.[33]

Alarmingly, the Copyright Office study decides that these requests are sufficient to strip a platform of its safe harbor.[34] It goes on to characterize users' anti-abuse proposals as attempts to "strip[] rightsholders from any realistic ability to enforce their (Congressionally mandated and constitutionally supported) rights."[35] In short, the Copyright Office's position as articulated in its Report is that all notices, no matter how obviously spurious or in bad faith, must be honored without further inquiry, and even the most cursory attempts at vetting will strip a platform of its safe harbor protections. This largely aligns with the position of rightsholders, who have balked at the idea of introducing additional safeguards into this system. Instead, these stakeholders insist

---

[32] *See, e.g.*, Caitlin Busch, *An Archive of Our Own: How AO3 Built a Nonprofit Fanfiction Empire and Safe Haven*, SyFy Wire (February 12, 2019) https://www.syfy.com/syfywire/an-archive-of-our-own-how-ao3-built-a-nonprofit-fanfiction-empire-and-safe-haven ; *See also The Digital Millennium Copyright Act at 22: What is it, why was it enacted, and where are we now? Before the Subcomm. on Intellectual Prop. of the S. Comm. on the Judiciary,* 116th Cong. 9 (2020) (Statement of Professor Rebecca Tushnet, Harvard Law School), https://www.judiciary.senate.gov/imo/media/doc/Tushnet%20Testimony.pdf (Archive of Our Own "receive[s] relatively few notices of claimed infringement, few of them are automated, and we subject them to individual review for validity. In the rare case that the notice complies with the DMCA and doesn't raise obvious fair use issues or assert non-copyright claims, our abuse team will remove the accused content and inform the user. Our experience with small-scale senders, consistent with the experience of many other OSPs, is that small-scale senders often consider DMCA claims to be a catch-all for objections such as that a work on the OTW's Archive has the same title as a different work they've published for sale or that they don't wish their name to be used in a work. Our experience with large-scale senders is that many are careful to avoid challenging non-exact copies, but unfortunately some do send takedown notices based on unhelpful metadata (e.g., title of a work even though the content is clearly different from that of the copyright claimant's work).").

[33] More than half of the DMCA takedown notices issued to Amazon's Kindle Direct, for example, are attempts to deliberately suppress a competitor's book from climbing the rankings. U.S. Copyright Office, *Section 512 Study: 9th Circuit Public Roundtable* 248 (May 13, 2016) (Testimony of Stephen Worth, Assoc. General Counsel of Amazon.com), https://www.copyright.gov/policy/section512/public-roundtable/transcript_05-13-2016.pdf ("[W]ith Kindle Direct publishing, authors routinely try to climb to the top spot in their category . . . by issuing bogus notices against higher ranking titles. And this for us actually accounts for more than half of the takedown notices that we receive.").

[34] USCO § 512 Report at 155.

[35] *Id*. at 169.

that targeted speech must be removed faster,[36] with a longer period before reinstatement,[37] and the removal must be executed without any human oversight or verification of claims.[38]

It bears repeating: The DMCA's notice-and-takedown provisions are extraordinarily powerful tools with a documented history of weaponization. It is true that artists face a difficult task in attempting to police the use of their copyrighted content online; however, we must acknowledge the enormous power of these takedown notices, their documented history of misuse, and the profound effect of that misuse on lawful speech. Asking for faster, more powerful notices with fewer safeguards is akin to discarding a tank and asking for a warhead.

## II. Algorithmic enforcement is not a viable answer

We cannot reasonably think about reforming Section 512 without understanding the private enforcement mechanisms that stakeholders have held out as possible solutions. Given the time and expense of federal litigation faced by rights-holders (and the pressures of operating at scale faced by platforms), it is unsurprising that many stakeholders have embraced the idea of technological solutions. But while automated private solutions "might sound good in theory," the messy realities of implementation -- technological limitations, complex legal protections and provisions, and the influence of a designer's commercial interests -- "raises a slew of questions regarding policy."[39] Private enforcement "can have the same far-reaching effect as actual law," including the ability to deprive users of legitimate income streams, "without any corresponding due process or accountability."[40] Because they operate automatically, these algorithms have the

[36] *Id.* at 159.
[37] *Id.* at 162.
[38] *Id.* at 152 n. 813.
[39] Lauren D. Shinn, *Youtube's Content ID as a Case Study of Private Copyright Enforcement Systems*, 43 AIPLA Q. J. 359, 372 (2015).

remarkable power to almost instantaneously erase speech -- including political speech, education, news, and speech which supports the livelihoods of millions of creators who derive their primary income via platforms with algorithmic content matching. A system which relies wholly on automated enforcement erases the very "safety valves" which prevent copyright law from becoming absolute, and violative of the First Amendment.

### A. Automated solutions are designed to answer the problems of their designers -- not anyone else.

The way in which these systems operate is determined by the particular needs, commercial interests, and resource limitations of the developer and any large stakeholders with which that developer is cooperating. Policymakers must grapple with what users and artists alike have understood for ages -- that the balance of equities in practice is determined less by the contours of law than by the aggregate results of numerous design choices which often have "more profitable" or "less profitable" answers, but rarely have clear right or wrong ones. An automated system which perfectly serves the needs of any one stakeholder -- be it platforms, commercial-scale rights-holders, users, or small artists -- invariably prejudices the interests of the remaining stakeholders.

### B. Automated solutions have multiple points of failure

We cannot reasonably think about reforming Section 512 without understanding the private enforcement mechanisms that stakeholders have held out as possible solutions. Given the time and expense of federal litigation faced by rights-holders (and the pressures of operating at scale faced by platforms), it is unsurprising that many stakeholders have embraced the idea of technological solutions. But while automated private solutions "might sound good in theory," the messy realities of implementation -- technological limitations, complex legal protections and

[40] *Id.*

provisions, and the influence of a designer's commercial interests -- "raises a slew of questions regarding policy."[41] Private enforcement "can have the same far-reaching effect as actual law," including the ability to deprive users of legitimate income streams, "without any corresponding due process or accountability."[42] Because they operate automatically, these algorithms have the remarkable power to almost instantaneously erase speech -- including political speech, education, news, and speech which supports the livelihoods of millions of creators who derive their primary income via platforms with algorithmic content matching.

Algorithmic matching has numerous steps, which we will necessarily simplify here. First, the system designer must compile and maintain a database of known content to which the algorithm can refer. A robust database contains, among other things, a reference file and ownership information for each work. The algorithm then uses reference files to create digital "fingerprints," which it compares against unknown media in an attempt to identify it.[43] When the algorithm returns a match, it provides rightsholders with a series of options. The scope and availability of these options depends on the design of the system, the level of access granted to the rights-holder, and other variables. Common options include claiming the content's ad revenue, taking the content offline (either *in toto* or selectively disabling the matching piece), or doing nothing.[44]

Though various kinds of errors can occur throughout this process, users are most frequently affected by "false positives" -- situations in which the algorithm incorrectly identifies content they have uploaded as infringing. Three common points of failure are errors in the database; erroneous flagging of content that does not match the reference file; and "content that

[41] Shinn at 372 (2015).
[42] *Id.*
[43] *See* David Kravets, *YouTube Alters Copyright Algorithms, Will 'Manually' Review Some Claims*, Wired (October 3, 2012) https://www.wired.com/2012/10/youtube-copyright-algorithm/.
[44] *See How Content ID Works: What Options Are Available to Copyright Owners?*, YouTube (2020), https://support.google.com/youtube/answer/2797370.

matches the reference file and is owned by the claimant, but constitutes a legal use of the content."[45]

*1. Database errors.*

The first category of false positives -- where the flagged content matches a reference file in the database, but the database's ownership information is incorrect -- can be broadly thought of as database errors. These happen for reasons that range from banal to malicious. Some database errors are caused by bad actors making false ownership claims, a problem that was particularly acute on YouTube in the early 2010s.[46] A low-quality or overbroad reference file can also cause an algorithm to throw false matches.[47] Selective additions of media to the database can also trigger improper takedowns, as when a new piece of media incorporates a pre-existing sample, and inclusion of the new media causes the algorithm to flag and remove the older clip.[48]

These kinds of false positives force us to confront difficult questions around database design, integrity, and access. In an ideal world, a content-matching database would be full of high-quality reference files, complete with thorough, current, and accurate information on ownership, licensing, and payment. An ideal database would also be widely open and available to artists who wish to use it to monitor (or monetize) their work. However, these two principles are often in tension; universal access creates a greater risk of introducing errors into the system, while curation creates gatekeeping power and an attendant risk of competitive concerns.

---

[45] Shinn at 372.

[46] Perhaps the most notable instance of this misuse was when a Russian group falsely claimed ownership over a number of viral cat videos, diverting the videos' ad revenue into their own pockets. David Kravets, *Rogues Falsely Claim Copyright on YouTube Videos to Hijack Ad Dollars*, Wired (November 21, 2011) https://www.wired.com/2011/11/youtube-filter-profiting/.

[47] *See* Urban Report at 90-92 (analyzing specific instances when targeted material did not match the allegedly infringed work).

[48] Notably, a 2016 episode of Family Guy "included a clip from 1980s Nintendo video game Double Dribble showing a glitch to get a free 3-point goal. Fox obtained the clip from YouTube where it had been sitting since it was first uploaded in 2009. Shortly after, Fox told YouTube the game footage infringed its copyrights. YouTube took it down." Fox dropped the claim and issued an apology when the story went viral. Andy, *Fox 'Stole' a Game Clip, Used it in Family Guy & DMCA'd the Original*, TorrentFreak (May 20, 2016) https://torrentfreak.com/fox-stole-a-game-clip-used-it-in-family-guy-dmcad-the-original-160520/.

As with algorithmic design more broadly, any commercial database will reflect the priorities of its designer. These influences affect who is allowed to populate the database, how that information is vetted or revised, the oversight and handling of ownership disputes, and the transparency (or lack thereof) regarding its operation. We need look no further than the debates surrounding YouTube's Content ID system to see the risks and trade-offs of a private, in-house fingerprinting system designed to address the business interests of a specific platform.[49]

*2. Strict vs "fuzzy" algorithms.*

The second failure case -- flagging content that does not match the reference file -- reflects yet another trade-off in algorithmic design. Algorithms that only flag exact or near-exact matches protect a greater range of unlicensed, yet legal, uses and exert less of a chilling influence on user speech. However, they are also easier to circumvent through basic manipulation of the underlying media, such as altering the tempo or pitch of a sound recording, or flipping a video to its mirror image.[50] Algorithms that flag "fuzzy" matches will be harder to evade, but will throw more false positives and stifle some legitimate uses of content.

It is worth noting that the degree of "fuzziness" in an algorithm is a design choice that explicitly prioritizes certain genres and styles of content over others. Fuzzy algorithms are good at catching and flagging algorithm-evading "edits" to popular content such as Top-40 hits. However, those same algorithms struggle when faced with classical and jazz music, where the underlying musical work is often in the public domain, and the difference between a copyrighted

---

[49] *See e.g.*, John Paul Titlow, *How YouTube Is Fixing Its Most Controversial Feature*, Fast Company (September 13, 2016) https://www.fastcompany.com/3062494/how-youtube-is-fixing-its-most-controversial-feature; Patrick McKay, *Open Letter to YouTube Regarding Content ID*, FairUseTube.org (September 15, 2011), http://fairusetube.org/articles/21-open-letter. *But cf.* SoundExchange Direct (2020), https://sxdirect.soundexchange.com/login/?next=/. While SoundExchange Direct is designed to organize metadata (rather than content fingerprinting for large-scale algorithmic enforcement), is a good example of how a database can be structured to accommodate the needs of artists. SoundExchange is, notably, a nonprofit.

[50] Nick Douglas, *You Can't Fool YouTube's Copyright Bots*, LifeHacker (January 24, 2018) https://lifehacker.com/you-cant-fool-youtubes-copyright-bots-1822174263.

recording and a public domain or live performance may be as little as a few notes on an improvisational section, or the sound qualify of the space in which it was recorded.[51] Content ID, often held up as the industry standard of content-matching, once erroneously flagged a video that was ten minutes of solid (original) white noise.[52]

*3. Legally permissible uses.*

Unlike an algorithm, copyright law is not binary or automated; the American system provides a number of exceptions and limitations that serve as a "safety valve" to protect legitimate policy ends. The Supreme Court has described these limitations and exceptions -- specifically citing fair use -- as "built-in First Amendment accommodations" to prevent copyright law from unduly burdening free speech.[53] These contours of copyright law, however, depend heavily on social, factual, and cultural context. The fundamental balance of copyright law rests in "[d]etailed doctrines … carefully designed to guide traditional, human law enforcement agents in addressing these questions" of appropriate unlicensed use.[54] Algorithmic enforcement, as a binary system designed to equate the *presence* of copyrighted content with its *misuse,* "is blatantly hostile to users' interests because it shifts the neutral presumption of fair use against them."[55] Moreover, systems such as Content ID allow rights-holders to instantaneously

[51] Michael Andor Brodeur, *Copyright bots and classical musicians are fighting online. The bots are winning.*, Wash. Post (May 21, 2020), https://www.washingtonpost.com/entertainment/music/copyright-bots-and-classical-musicians-are-fighting-online-the-bots-are-winning/2020/05/20/a11e349c-98ae-11ea-89fd-28fb313d1886_story.html. *See also* Ulrich Kaiser, *Can Beethoven Send Takedown Requests? A First-hand Account of One German Professor's Experience With Overly Broad Upload Filters*, Wikimedia Found. (August 27, 2018) https://wikimediafoundation.org/news/2018/08/27/can-beethoven-send-takedown-requests-a-first-hand-account-of-one-german-professors-experience-with-overly-broad-upload-filters/.

[52] Chris Baraniuk, *White Noise Video on YouTube Hit by Five Copyright Claims*, BBC News (January 5, 2018) https://www.bbc.com/news/technology-42580523; Timothy Geigner, *White Noise on YouTube Gets FIVE Separate Copyright Claims From Other White Noise Providers*, TechDirt (January 5, 2018) https://www.techdirt.com/articles/20180105/10292038938/white-noise-youtube-gets-five-separate-copyright-claims-other-white-noise-providers.shtml

[53] *Eldred v. Ashcroft,* 537 U.S. 186, 219 (2003).

[54] Maayan Perel and Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 Stan. Tech. L. Rev. 473 (2016), https://www-cdn.law.stanford.edu/wp-content/uploads/2016/10/Accountability-in-Algorithmic-Copyright-Enforcement.pdf.

divert revenue streams away from claimees upon filing a claim, leading to lost or delayed revenue, as well as a host of secondary knock-on effects for the user whose speech has been removed.[56]

## Conclusion

Two hundred and twenty-nine million American adults live their lives online under the shadow cast by Section 512. Whatever the risks or rewards, we cannot be reckless with the speech rights of those who find themselves governed by the system we create. Congress must acknowledge that this debate is not happening in a vacuum, and reject the fantasy of copyright being a struggle between "tech" and "content." Copyright law, broadband access, algorithmic governance, and economic incentive structures are all intertwined, and all impact Americans' ability to speak online. In a moment of massive social change, we must not take that for granted.

[55] Taylor B. Bartholomew, *The Death of Fair Use in Cyberspace: YouTube and the Problem with Content ID*, 13 Duke L. & Tech. Rev. 66, 68 (2015), (https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1271&context=dltr.

[56] Amanda Perelli, *Prominent Youtube Creator Lindsay Ellis Is Challenging the Platform Over the Way it Handles Copyright Claims*, Bus. Insider, (Oct. 29, 2019) https://www.businessinsider.com/youtuber-lindsay-ellis-fights-platform-universal-over-copyright-claim-2019-10.