Indiana Law Review

NOTES

THE FUTURE OF THE DIGITAL MILLENNIUM COPYRIGHT ACT: HOW AUTOMATION AND CROWDSOURCING CAN PROTECT FAIR USE

ADAM EAKMAN[*]

INTRODUCTION

On August 6, 2012, NASA completed one of the most technologically advanced missions in the agency's history.[1] After a decade of work and input from hundreds of employees, NASA successfully landed a robotic rover named Curiosity on the planet Mars.[2] Curiosity travelled over 350 million miles and landed almost flawlessly.[3] NASA anticipated a global audience for this historic event and prepared a YouTube channel to exhibit video taken from this mission.[4] Just one hour after Curiosity landed on Mars, while popping champagne and exchanging high-fives, NASA engineers posted a thirteen-minute video of the extraordinary landing on their YouTube channel.[5] Ten minutes later the video was inaccessible.[6]

Anyone who tried to access the video encountered this message: "This video contains content from Scripps Local News, who has blocked it on copyright grounds."[7] This video was created by NASA and was posted on NASA's official YouTube channel, yet was ultimately removed by a private news service.[8] This video did not contain any copyright-infringing material; the removal was simply

---

1. *NASA Facts: Mars Science Laboratory/Curiosity*, NAT'L AERONAUTICS & SPACE ADMIN., http://www.jpl.nasa.gov/news/fact_sheets/mars-science-laboratory.pdf (last visited Feb. 21, 2014), *archived at* http://perma.cc/ZSV4-M8GB.

2. *Id.*

3. Jason Hanna, *'Impressive' Curiosity Landing Only 1.5 Miles Off, NASA Says*, CNN (Aug. 14, 2012, 6:25 PM), http://www.cnn.com/2012/08/10/us/mars-curiosity/index.html?eref=mrss_igoogle_cnn, *archived at* http://perma.cc/3KSB-A4GU.

4. Alex Pasternack, *NASA's Mars Rover Crashed into a DMCA Takedown*, MOTHERBOARD (Aug. 6, 2012, 11:49 AM), http://motherboard.vice.com/blog/nasa-s-mars-rover-crashed-into-a-dmca-takedown, *archived at* http://perma.cc/A2GH-ZDLU.

5. *Id.*

6. *Id.*

7. *Id.*

8. *Id.*

a mistake.[9] Nor was this the first time such a mistake happened.[10] Just four months earlier, NASA posted a video on YouTube of the Space Shuttle *Discovery* being transported to the Kennedy Space Center that was also inadvertently removed.[11] In fact, a NASA official stated that videos on NASA's YouTube channel are mistakenly removed about once per month.[12]

Yet why would someone make such an obvious error and mistake footage of Mars' terrain for copyright infringement? The answer is no one did. No *person* claimed that this video infringed on a copyright. Rather, an automated process designed to find and eliminate copyright infringement generated a request to YouTube to remove this material, mistaking it for genuine piracy.[13] In turn, YouTube's own automated process responded to this request and removed the material quickly.[14] The video was posted, identified, and removed in less than ten minutes.[15]

This process was part of a routine application of the procedure prescribed in the Digital Millennium Copyright Act ("DMCA").[16] At the close of the twentieth century, Congress passed the DMCA in an attempt to protect copyright-protected material from Internet piracy while still allowing the growth of creative expression on the Internet.[17] By providing "safe harbors" for websites that may host third-party content and meet the law's prescribed conditions, the DMCA has allowed for the exponential growth of websites such as Google, YouTube, Facebook, and Amazon without the threat of constant copyright-related litigation.[18] By following the procedures required under the DMCA, a website is able to respond to copyright infringement requests quickly, easily, and with confidence that it will not face litigation.[19]

Today, application of the DMCA has grown as quickly as the websites it is designed to protect.[20] Although it is impossible to know the number of DMCA requests that are generated each day, the few corporations that publish the number of DMCA requests they review provide staggering numbers.[21] By the close of

---

9. *Id.*

10. *Id.*

11. Joshua Kostein, *NASA's New Rover Arrives on Mars, Crash Lands in YouTube's DMCA Hell*, VERGE (Aug. 6, 2012, 5:24 PM), http://www.theverge.com/2012/8/6/3223820/nasa-rover-youtube-copyright-takedown, *archived at* http://perma.cc/Z9RZ-XN93.

12. *Id.*

13. Pasternack, *supra* note 4.

14. *Id.*

15. *Id.*

16. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512 (2013).

17. S. REP. NO. 105-190, at 2 (1998).

18. Edward Lee, *Decoding the DMCA Safe Harbors*, 32 COLUM. J.L. & ARTS 233, 259-60 (2009).

19. *Id.*

20. *See Google Transparency Report*, GOOGLE, http://www.google.com/transparencyreport/removals/copyright/ (last visited January 9, 2014), *archived at* http://perma.cc/AJW3-5E8J.

21. *See, e.g.*, *id.*

2013, Google was receiving twenty-five million DMCA takedown requests per month.[22] For many Internet-based businesses like Google that receive a flood of DMCA takedown requests daily, the Herculean task of responding to all of these requests has become overwhelming.[23]

For these businesses the only viable solution is automation.[24] For large companies like Microsoft, the task of managing DMCA requests for their vast array of copyrights has been entrusted to outside firms which use an automated process to efficiently generate DMCA takedown requests for their clients.[25] Degban, for example, is a company that specializes in helping other companies combat piracy of their copyrighted material using "innovative intelligent technology."[26] The sheer number of DMCA requests companies like Degban generate leave little doubt that automation is involved in this process.[27] Degban alone has generated DMCA requests to remove over one hundred-twenty million URL's from Google since March 2011.[28]

Other websites that host third-party material use an in-house automated process to remove potentially infringing material.[29] YouTube, a website which allows third parties to upload and watch videos, employs an automated system called Content ID for detecting and removing potentially infringing material.[30] If a user uploads a video that matches a database of copyright-protected videos managed by Content ID, the user's video may be removed by the copyright holder in a matter of minutes.[31] While a user can fight a wrongful takedown, this process can take up to a month.[32]

These structures provide a glimpse into a future in which copyright-protected material is monitored exclusively by automated systems.[33] In the future, it is possible that automated systems, rather than judicious people, will be both

---

22. *Id.*

23. *See* Tim Cushing, *LeakID and the DMCA Takedown Notice Farce*, Techdirt (Nov. 2, 2012, 8:32 AM), http://www.techdirt.com/articles/20121026/17190420859/leakid-dmca-takedown-notice-farce.shtml, *archived at* http://perma.cc/BAY6-DWPK.

24. *Id.*

25. *Id.*

26. *About Us*, Degban, http://www.degban.com/aboutus.html (last visited Oct. 3, 2014), *archived at* http://perma.cc/B756-2U8R.

27. Cushing, *supra* note 23.

28. *Google Transparency Report*, *supra* note 20.

29. *See How Content ID Works*, YouTube, http://www.youtube.com/t/contentid (last visited Oct. 3, 2014), *archived at* http://perma.cc/XH4S-TKXG.

30. *Id.*

31. *Id.*

32. Paul Tassi, *The Injustice of YouTube ContentID Crackdown Reveals Google's Dark Side*, Forbes (Dec. 19, 2013, 10:00 AM), http://www.forbes.com/sites/insertcoin/2013/12/19/the-injustice-of-the-youtube-content-id-crackdown-reveals-googles-dark-side/, *archived at* http://perma.cc/XP9K-JMY8.

33. *See infra* Part III.

generating and responding to requests to remove information.[34] This could mean that the information available through the Internet may not be limited by any human creativity, but rather, by automated processes designed to protect copyright-protected material.[35] Computers, rather than people, would determine the future composition of the Internet.[36]

Internet-based businesses that must scour all the content they host and separate copyright-infringing material from legitimate speech face several questions that this article seeks to address. First, is automation an effective solution to this problem? More specifically, are computer programs capable of accurately detecting when copyright-protected material is being pirated on the Internet while still protecting non-infringing material? Second, if automation is not capable of effectively identifying copyright infringement, what should these businesses do to protect both copyrighted material and non-infringing material? With the number of takedown requests to remove material under the DMCA constantly growing, Internet-based businesses need a realistic way to respond to these requests quickly and easily.[37] If an automated process is capable of detecting and removing copyright-protected material, then this process is capable of saving thousands of hours of work and preventing hundreds of lawsuits.[38] However, if an automated process cannot accurately detect genuine piracy, then non-infringing speech may be suppressed.[39] For businesses that seek to protect non-infringing material, yet want to avoid any possible liability, there is a strong demand for any solution that can solve these twin goals. This Note seeks to provide a solution that satisfies these seemingly contradictory goals.

Part I of this Note provides an overview of the Digital Millennium Copyright Act including the purpose of its creation. It explores what the text of the DMCA requires of copyright holders and third parties hosting copyright-protected information. Part II examines how the DMCA is being applied today. More specifically, it examines how application of the DMCA has increasingly required the use of an automated process in order for large Internet-based companies to comply with the requirements of the DMCA. Part III explores the legal implications of employing an automated process for filing and responding to DMCA requests. This Note argues that a totally automated process for detecting and removing copyright-protected material from the Internet does not adequately protect original works that do not violate any copyrights because such use is considered "fair use." Because fully-automated systems cannot totally distinguish piracy from fair use, relying solely on automated systems will result in the removal of non-infringing material and may even lead to legal liabilities under the DMCA. Finally, Part IV suggests that the solution to this quagmire lies

---

34. *Id.*

35. *Id.*

36. *Id.*

37. *See* Cushing, *supra* note 23.

38. *See* Lee, *supra* note 18.

39. *See generally* Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment*, 24 HARV. J.L. & TECH. 171 (2010).

in "crowdsourcing" the task. Crowdsourcing refers to allowing a broad online community of users to complete a task, and in this case, it is a viable solution to separating copyright-protected material from piracy. By asking the users themselves to identify the nature of hosted content, a business shifts the task of separating the wheat from the chaff away from an automated process to millions of discerning people. While an automated process can be used to eliminate blatant copyright violations efficiently, human input is still required in ambiguous cases. Crowdsourcing would thus allow businesses to collect input from many citizens as to whether the material is fair use. Ultimately, crowdsourcing offers a solution that is affordable, efficient, and accurate. Under this model, Internet-based businesses would be able to comply with requests to remove protected information in accordance with the DMCA and do so with increased confidence that they are not removing non-infringing material.

I. UNDERSTANDING THE DIGITAL MILLENNIUM COPYRIGHT ACT

At the close of the twentieth century the World Wide Web was still in its infancy and legislators sought to protect this fledgling industry from piracy and aid in its growth.[40] The DMCA was passed in 1998 with the purpose of promoting "the robust development and world-wide expansion of electronic commerce, communications, research, development, and education in the digital age."[41] With this law, the legislature hoped to bring United States copyright law "squarely into the digital age" and offer greater protections to copyright holders who host their material over the Internet.[42] Furthermore, the DMCA sought to increase Internet access to "the movies, music, software, and literary works that are the fruit of American creative genius."[43] Ultimately, the DMCA strove to strike a delicate balance between fighting piracy quickly and effectively, while refraining from trampling on fair use copyright-protected material.[44]

A. The Requirements of the DMCA

The DMCA outlines specific step-by-step procedures that must be followed by copyright holders and websites that host copyrighted material anytime infringing material is found.[45] If these steps are followed correctly, copyright holders can be confident that piracy will be removed quickly while third parties

---

40. S. REP. NO. 105-190, at 1-2 (1998).

41. *Id.*

42. *Id.* at 2.

43. *Id.*

44. *See* Thomson Reuters, *Drafting DMCA Takedown Notices*, PRACTICAL LAW (Apr. 2, 2013), http://us.practicallaw.com/8-525-5186?q=&qp=&qo=&qe=#null, *archived at* http://perma.cc/ BH3P-PQRV.

45. Richard Chapo, *DMCA Process—From Agent Designation to Takedown Notices*, SOCALINTERNETLAWYER.COM (July 12, 2012), http://www.socalinternetlawyer.com/dmca-process/, *archived at* http://perma.cc/CJ3D-BV3T.

that hosted that material can be equally confident they will not face litigation.[46] In addition, the law provides unique legal protections to Internet service providers ("ISPs") by extending them "safe harbors" if they follow specific steps when hosting third-party content.[47] The statute defined an ISP as "an entity offering the transmission, routing, or providing of connections for digital online communications, between or among points specified by a user, of material of the user's choosing, without modification to the content of the material as sent or received."[48] If an ISP falls under the statute's safe harbor provision, the ISP will not be liable for hosting pirated material on its website.[49]

Needless to say, this creates a strong incentive for all ISPs to be certain they are legally protected by the DMCA's safe harbor protections.[50] In order to qualify for safe harbor protection the ISP must have no actual knowledge or awareness of facts or circumstances that would make it apparent that the material on its system or network is infringing on copyright-protected material.[51] Once the ISP becomes aware that it is hosting infringing material, it must act "expeditiously" to remove access to the material.[52] In addition to always following these procedures, the ISP cannot receive a direct "financial benefit" from the infringing material,[53] must maintain a designated agent who may be contacted by copyright holders in the event of infringement,[54] and must have a policy for terminating repeat offenders.[55]

Today, virtually all ISPs that host any third-party material must make sure they comply with these DMCA safe harbor provisions in order to insulate themselves from liability.[56] In fact, "[e]very internet company in the United States that deals with content of third-party users—companies such as Amazon, AOL, CNN, eBay, Facebook, Google, MySpace, YouTube, and numerous startups aspiring to become just as successful—must adopt and implement a DMCA policy to fall within the safe harbors."[57] Failing to do so may even be considered a breach of corporate duty.[58]

---

46. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512 (2013).

47. *Id.* § 512(c).

48. *Id.* § 512(k).

49. *Id.* § 512(a).

50. STEPHEN FISHMAN, THE COPYRIGHT HANDBOOK: WHAT EVERY WRITER NEEDS TO KNOW 319 (Ilona Bray ed., 11th ed. 2011).

51. 17 U.S.C. § 512(c)(1)(A) (2013).

52. *Id.*

53. *Id.* § 512(c)(1)(B).

54. *Id.* § 512(c)(2).

55. *Id.* § 512(i)(l)(A).

56. Lee, *supra* note 18, at 233-34.

57. *Id.*

58. *Id.* at 234.

B. Notice and Takedown Procedures

The DMCA envisioned that most ISPs would be protected from lawsuits for hosting copyright-protected material through the law's safe harbor provisions.[59] By protecting ISPs, however, the law placed the burden on copyright holders to find cases of infringement.[60] To compensate for this burden shifting, the DMCA provides a specific procedure that, if successfully followed by the copyright holder, will result in the expeditious elimination of pirated material.[61]

First, when a copyright holder finds his or her material is being pirated, the copyright holder must file a notice with the ISP hosting the material and begin the "notice and takedown procedure."[62] As long as this notice conforms to the standards required by the statute, the ISP must act "expeditiously" to remove or block access to the infringing material as soon as the notice is received.[63] The ISP must then provide the party that posted the contested material with a notice stating that it has been taken down.[64] If the party that had his or her material taken down believes this was an error and the material does not violate any copyrights, the party can file a counter notification.[65] After a counter notification is filed, the originator of the first DMCA takedown notice is then informed that a counter notification has been filed, and the material will be restored in ten to fourteen business days unless the originator seeks a court order to restrain the posting of the material.[66] If, on the other hand, a counter notification is never filed, the material may be permanently removed.[67]

These procedures were created in order to streamline the identification and removal of pirated material with little fanfare and rare litigation.[68] They create a simple recipe that will almost certainly result in the removal of the material.[69] DMCA takedown requests can be filed in a matter of minutes.[70] One website,

---

59. *See generally* Diane M. Barker, *Defining the Contours of the Digital Millennium Copyright Act: The Growing Body of Case Law Surrounding the DMCA*, 20 BERKELEY TECH. L.J. 47, 57 (2005) (explaining that the DMCA safe-harbor provisions were intended to provide ISPs immunity from the infringing behavior of their subscribers).

60. *Id.* at 58.

61. 17 U.S.C. § 512(c) (2013).

62. *Id.*

63. *Id.* § 512(c)(3).

64. *Id.* § 512(g).

65. *Id.* § 512(g)(3).

66. *Id.* § 512(g)(2).

67. *Id.*

68. *See* Brian J. Meli, *Four Letters No Digital Content Provider Should Ever Forget: DMCA*, LEGALMATTERBLOG.COM (Oct. 31, 2013), http://legalmatterblog.com/2013/10/31/four-letters-no-digital-content-provider-should-ever-forget-dmca/, *archived at* http://perma.cc/5T5H-CLV2 (stating that the DMCA's purpose was to streamline takedowns while protecting service providers from litigation).

69. *Id.*

70. *How Can I File a DMCA Takedown Notice?*, DMCA.COM, http://www.dmca.com/FAQ/

Regainyourname.com, will even draft a DMCA takedown request for its users.[71] The only information needed is your name, address, email, a link to the infringing URLs, the name of your business, and the search terms used to find the website, and you can send your own DMCA request in seconds.[72] Such a simple procedure for filing a DMCA takedown request has produced an unintended side effect.[73] By promoting the rapid generation of takedown requests, and requiring an expeditious response to these requests, the DMCA has essentially promoted an environment that favors speed over accuracy.[74] ISPs are incentivized to respond to these requests without examining their validity.[75] By emphasizing speed of removal, the DMCA has made automation of DMCA takedown request nearly inevitable.[76]

II. DMCA TAKEDOWN REQUESTS GENERATED THROUGH AUTOMATION

The limitations of this carefully prescribed process can only be fully understood through a thorough examination of how the DMCA takedown procedures are being applied today. Microsoft is a powerful example. Microsoft holds thousands of copyrights and has the daunting task of ensuring that no one violates any of their copyrights on the Internet.[77] Rather than expend valuable resources on this interminable task, Microsoft hired an outside firm to identify piracy and generate DMCA takedown requests.[78] In order to find every case of possible infringement, this company employed an automated process that used keyword searches to find possibly infringing material, and then generated requests based on the likelihood that the search returned pirated material.[79] This

---

How-can-I-file-a-DMCA-Takedown-Notice (last visited Oct. 3, 2014), *archived at* http://perma.cc/ 43G7-88MZ.

71. *DMCA Copyright Infringement Take-Down Drafting Service*, REGAIN YOUR NAME, http://regainyourname.com/dmca-copyright-takedown-notices-for-stolen-material/ (last visited Oct. 3, 2014), *archived at* http://perma.cc/4EHN-ZUFX.

72. *Id.*

73. *See infra* Part II.

74. *Id.*

75. *Id.*

76. *Id.*

77. *Legal Resources*, MICROSOFT, http://www.microsoft.com/en-us/legal/Intellectual Property/Permissions/Default.aspx (last visited Nov. 27, 2014), *archived at* http://perma.cc/3Q32-JYXH (describing the numerous copyrighted material Microsoft owns).

78. Cushing, *supra* note 23. However, Microsoft fired LeakID from managing their DMCA request just hours after TorrentFreak discovered that the firm generated another embarrassing list of DMCA takedown requests that included the Wikipedia entry on Microsoft. *See* Ernesto, *Microsoft Ditches Anti-Piracy Partner After Embarrassing DMCA Takedowns*, TORRENTFREAK (Sep. 27, 2013), http://torrentfreak.com/microsoft-ditches-anti-piracy-partner-after-embarrassing-dmca-takedowns-130927/, *archived at* http://perma.cc/AA9W-F2KP.

79. Cushing, *supra* note 23 (referring to how LeakID uses a "patrolbot" to find copyright infringement).

process had little human oversight and was prone to error.[80] In fact, in 2013, Google received a request from Microsoft to takedown portions of their own website, Microsoft.com, claiming they infringed on their own copyrights.[81] While this approach certainly has errors, it is through this automated process that Microsoft was able to generate DMCA takedown requests to remove over ten million URLs from Google in 2012 alone.[82]

Microsoft is not the only corporation generating ridiculous DMCA takedown requests. In April 2012, Warner Brothers was releasing the hopeful blockbuster movie *Wrath of the Titans*, yet they thwarted their own attempts at creating publicity about the film by generating DMCA takedown requests that would remove valuable information about the film.[83] Warner Brothers inadvertently asked Google to remove links to the Internet Movie Database listing of *Wrath of the Titans*, links to the official trailer on Apple and Hulu, a review of the movie posted on BBC America, and even a listing that helped people who wanted to watch the movie find theaters near them where it was going to be playing.[84] The Recording Industry of America ("RIAA") also shot itself in the foot when it filed a takedown request to remove publicity about one of its own recording artists, Lady Antebellum.[85] Not only did the RIAA request that *The Guardian* remove a review of a Lady Antebellum album, but also demanded that Google remove a link to reviews of Lady Antebellum songs hosted by AOL's music site.[86]

In response to many copyright holders' blitz of DMCA requests, ISPs have started implementing automated programs that preemptively remove content the program detects as infringing before even receiving a DMCA request.[87] YouTube's "Content ID" scans all videos that are uploaded to its site and looks

---

80. *Id.*

81. Ernesto, *Microsoft Wants Google to Censor . . . Microsoft.com*, TORRENTFREAK (July 28, 2013), http://torrentfreak.com/microsoft-wants-google-to-censor-microsoft-com-130728/, *archived at* http://perma.cc/J4L6-5XEF.

82. *Google Transparency Report*, *supra* note 20.

83. *Video DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Apr. 17, 2012), http://www.chillingeffects.org/notice.cgi?sID=291695, *archived at* http://perma.cc/KUN6-G2ZH; *see also* Enigmax, *Copyright Holders Punish Themselves with Crazy DMCA Takedowns*, TORRENTFREAK (May 25, 2012), http://torrentfreak.com/copyright-holders-punish-themselves-with-crazy-dmca-takedowns-120525/, *archived at* http://perma.cc/N9FT-QKPM (stating that "Warner and their anti-piracy partners managed to undermine their own marketing campaign for *Wrath of the Titans* with DMCAs sent to Google.").

84. Enigmax, *supra* note 83.

85. Mike Masnick, *DMCA Notices So Stupid It Hurts*, TECHDIRT (May 30, 2012), http://www.techdirt.com/articles/20120527/23520719089/dmca-notices-so-stupid-it-hurts.shtml, *archived at* http://perma.cc/94FZ-VPUS.

86. *Id.*

87. *A Guide to YouTube Removals: So My Video Was Removed from YouTube . . . What Do I Need To Know?*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/ru/issues/intellectual-property/guide-to-youtube-removals (last visited Oct. 3, 2014), *archived at* http://perma.cc/8NW4-6E94.

for any matches with a reference database of copyright-protected material that copyright owners have provided.[88] If Content ID detects a match, the copyright holder is notified and given the opportunity to remove the material.[89] Every day, Content ID scans 400 years of video, and so far has detected infringement in over 200 million videos.[90] This system works quickly, efficiently, and has greatly diminished the amount of pirated content available on YouTube, but it remains an automated system with known flaws.[91] In the fall of 2012, YouTube was live-streaming the Democratic National Convention and just before Michelle Obama began her speech, the video was blocked because in the background a copyright-protected song was playing.[92] This song was played throughout the campaign, and the Democratic National Convention almost certainly had rights to play this song,[93] but the video was blocked nonetheless.[94] A similar mistake was made when YouTube was live streaming the Hugo Awards.[95] The entire stream was blocked right before screenwriter Neil Gaiman accepted an award for his work on an episode of *Dr. Who* because the ceremony played a clip of the episode.[96] Content ID is simply unable to distinguish when an individual may have legal rights to use copyright-protected material.[97]

These automated programs used for detecting copyright infringement have their demonstrable flaws, but at the rate DMCA requests have been growing, automation may seem like the only way ISPs could possibly keep up. In the month of August 2013 alone, Google received DMCA requests to remove over twenty million URLS.[98] By the close of 2013, Google was on track to receive requests to remove more than 235 million links,[99] which is more than quadruple

88. *Id.*

89. *Id*.

90. *Statistics*, YOUTUBE, http://www.youtube.com/yt/press/statistics.html (last visited Feb. 21, 2014), *archived at* http://perma.cc/783Z-7QQ7.

91. *See, e.g.*, Adam Holland, *Robots Mistakenly Take Down Livestreams*, CHILLING EFFECTS (Sep. 18, 2012), https://www.chillingeffects.org/weather.cgi?WeatherID=658, *archived at* http://perma.cc/FR5R-CFN7; Zachary Knight, *Copyright Enforcement Bots Seek and Destroy Hugo Awards*, TECHDIRT (Sep. 4, 2012), http://www.techdirt.com/articles/20120903/18505820259/ copyright-enforcement-bots-seek-destroy-hugo-awards.shtml, *archived at* http://perma.cc/U4UW-2N3F.

92. *Id.*

93. Tim Cushing, *Copyright Killbots Strike Again: Official DNC Livestream Taken Down By Just About Every Copyright Holder*, TECHDIRT (Sep. 5, 2012), http://www.techdirt. com/articles/20120904/22172920275/copyright-killbots-strike-again-official-dnc-livestream-taken-down-just-about-every-copyright-holder.shtml, *archived at* http://perma.cc/G5GY-3URJ (calling the stream "perfectly legal").

94. Holland, *supra* note 91.

95. Knight, *supra* note 91.

96. *Id.*

97. Holland, *supra* note 91.

98. *Google Transparency Report*, *supra* note 20.

99. Ernesto, *Google Discarded 21,000,000 Takedown Requests in 2013*, TORRENTFREAK

the amount it received just one year earlier.[100] In fact, Google is deleting links to "pirated" material at a rate of nine links per second.[101] If Google hired lawyers to personally review all of the requests it received during the week of September 9, 2013, and each of these lawyers worked without breaks eight hours a day, spending only thirty seconds reviewing each URL removal request, Google would still need to hire over 700 lawyers devoted solely to reviewing these requests.[102] The growth of the DMCA notice and takedown procedures through automation has become so rampant that for many businesses, automation has become a necessity before the legal ramifications can be thoroughly considered.[103]

While automated programs are very effective at detecting wholesale copyright infringement, the greatest stumbling block for these programs is when material is protected by the fair use doctrine. At its core, the fair use doctrine allows individuals to legally use material that would otherwise be copyright-protected material.[104] As the United States Supreme Court stated, "anyone . . . who makes a fair use of the work is not an infringer of the copyright with respect to such use."[105] Under a case-by-case analysis, courts use four factors to determine whether an individual who used copyright-protected material is not liable because the use was considered a fair use.[106] These factors are:

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and 4) the effect of the use upon the potential market for or value of the copyrighted work.[107]

If the fair use doctrine applies, an individual can use copyright-protected material without permission and remain immune from litigation.[108] It is this doctrine that allows movie reviewers to show clips of films, musicians to parody popular

---

(Dec. 27, 2013), http://torrentfreak.com/google-discarded-21000000-takedown-requests-in-2013-131227/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Torrentfreak +(Torrentfreak), *archived at* http://perma.cc/LQ8J-CREX.

100. Zoe Fox, *Google Copyright Infringement Reports to Quadruple This Year*, MASHABLE (July 31, 2013), http://mashable.com/2013/07/31/google-copyright-reports/, *archived at* http://perma.cc/7KU5-CYN3.

101. Eric Limer, *Google Is Now Deleting Nine "Pirate" Links Every Second*, GIZMODO (Oct. 5, 2013), http://gizmodo.com/google-is-now-deleting-nine-pirate-links-every-second-1441566720, *archived at* http://perma.cc/YA3C-5W9K.

102. *Google Transparency Report*, *supra* note 20.

103. *See* Fox, *supra* note 100.

104. Neil Weinstock Netanel, *Making Sense of Fair Use*, 15 LEWIS & CLARK L. REV. 715, 719 (2011).

105. Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417, 433 (1984).

106. 17 U.S.C. § 107 (2013).

107. *Id.*

108. Netanel, *supra* note 104, at 10.

songs, and teachers to show television shows to their students without ever obtaining permission from the copyright holder.[109]

Because the fair use doctrine allows for the authorized use of copyright-protected material, experience has shown that Content ID, and other automated programs simply do not adequately account for fair use.[110] For example, in 2009 Jonathan McIntosh remixed scenes from the popular *Twilight* movies with clips from *Buffy the Vampire Slayer* in his video *Buffy v. Edward: Twilight Remixed* in order to critique the misogynistic portrayals of females in the *Twilight* films.[111] This was such a clear case of fair use that the United States Copyright Office cited it in their own report as an example of "transformative noncommercial" work and an exemplar of fair use.[112] Nonetheless, the video was removed from YouTube after Lionsgate Films bought Summit Entertainment and asked that the video be removed.[113]

Fair use regarding political speech presents even more worrisome problems. In order to promote his 2008 presidential campaign, John McCain created a YouTube channel to post his political ads and spread publicity.[114] Just weeks before the election, several of John McCain's videos were removed, citing copyright infringement.[115] These videos did contain copyrighted materials but clearly fell under the fair use exception because they were part of a political campaign.[116] After YouTube was notified of this, the McCain campaign was still forced to wait more than a week for the videos to be restored while losing critical time at the absolute height of the election season.[117] As Wendy Seltzer, a staff attorney with the Electronic Frontier Foundation and Fellow with Harvard's Berkman Center for Internet and Society stated, "[i]f there was ever a clear case of non-infringing fair use—speech protected by the First Amendment—this should have been it: a political candidate, seeking to engage in public multimedia

---

109.   Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems*, 15 HARV. J.L. & TECH. 41, 43 (2001).

110.   *See* Fred Von Lohmann, *YouTube's January Fair Use Massacre*, ELECTRONIC FRONTIER FOUNDATION (Feb. 3, 2009), https://www.eff.org/deeplinks/2009/01/youtubes-january-fair-use-massacre, *archived at* http://perma.cc/AU69-Q4D2 (written in response to YouTube removing a video of a young teenager singing "Winter Wonderland.").

111.   Adam Holland, *Takedown of the Week: YouTube and Lionsgate Films Continue to Ban Paradigmatic Example of Fair Use*, CHILLING EFFECTS (Jan. 10, 2013), http://cached. newslookup.com/cached.php?ref_id=429&siteid=2350&id=3833891&t=1385439475, *archived at* http://perma.cc/4MPN-7P49*; see also* Daniel Nye Griffiths, *Copyright in the Twilight Zone: The Strange Case of 'Buffy Versus Edward,'* FORBES (Jan. 15, 2013), http://www.forbes.com/sites/ danielnyegriffiths/2013/01/15/copyright-in-the-twilight-zone-the-strange-case-of-buffy-versus-edward/, *archived at* http://perma.cc/6ZZM-XN9C.

112.   Holland, *supra* note 111.

113.   *Id.*

114.   Seltzer, *supra* note 39, at 172.

115.   *Id.*

116.   *Id.* at 172-73.

117.   *Id.*

debate, used video snippets from the television programs on which the issues were discussed."[118]   Despite the obvious fair use application, YouTube's automated system removed the material.[119]

Examples like these illustrate that the stakes regarding DMCA takedown requests are high. ISPs that host third party material do not want to become havens of pirated material, but they also want to protect their users who may want to upload their own artistic expressions. Automation can be a powerful tool for fighting piracy, but a tool that should be employed with caution. Experience has shown that automation, when not given proper oversight, will trample on legitimate fair use of content. This Note seeks to provide legal guidance to businesses that would like to generate and respond to DMCA takedown requests quickly and efficiently, yet still protect fair use.

III. LEGAL ANALYSIS

Automated generation and response to DMCA requests seems inevitable, but before a company chooses to pursue this route it must first consider the legal liabilities that may accompany automation. Can an automated process designed to scour the Internet looking for piracy accurately identify cases of genuine copyright infringement? If not, what are the legal consequences for businesses that use an automated process with known flaws?

Employing an automated process for detecting illegal activity is a growing trend in the United States and has proven to be a valuable resource in fighting illegal activities.[120]   Many cities throughout the country employ "red-light cameras" which use cameras, magnetic fields, and a small computer to detect when someone runs a red light.[121]   In fact, many of these programs can collect the date, time of the infraction, the location, the speed of the car as it ran the red light, and the license plate number of the vehicle.[122] This allows tickets to be sent to the drivers without any police intervention whatsoever.[123] Automation can also be used for more complex violations.[124] In July 2013, the Securities and Exchange Commission announced they would be using an automated process to scour

---

118. *Id.* at 173.

119. *Id*. at 172.

120. *See, e.g.*, Tom Harris, *How Red-Light Cameras Work*, HOW STUFF WORKS, http://auto.howstuffworks.com/car-driving-safety/safety-regulatory-devices/red-light-camera.htm (last visited Feb. 21, 2014), *archived at* http://perma.cc/2YCW-EFLA; Press Release, U.S. Securities and Exchange Commission, SEC Announces Enforcement Initiatives to Combat Financial Reporting and Microcap Fraud and Enhance Risk Analysis, *available at* http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171624975#.UkQhxIasim5, *archived at* http://perma.cc/RA96-T8Z8.

121. Harris, *supra* note 120.

122. *Id.*

123. *Id.*

124. *Id.*

corporate filings looking for signs of accounting fraud.[125] This program is called "Robocop" and the SEC plans to expand its capabilities.[126]

However, not all aspects of the law lend themselves to automation. Some legal concepts are not defined clearly enough to allow computer programs to adequately detect when a violation occurs. For example, it would likely be impossible to create a computer program to scour the Internet looking for "hard-core pornography" because there is no bright-line test for determining which depictions of sexual acts are art and which are hard-core pornography.[127] Justice Stewart's candid statement that he cannot precisely define hard-core pornography, "[b]ut I know it when I see it,"[128] simply cannot be converted to an algorithm for an automated process. Without a bright-line test, an automated process has no rule it can apply to determine when there are violations of that rule.[129]

These examples provide two ends of a continuum in which copyright infringement can be compared. Is piracy like running a red light; an act that can be easily detected through automation? Or is piracy more like identifying pornography; something that cannot be simplified to a test and is subject to significant variation? Experience has shown that the answer is both.

In many ways, piracy is like running a red light, and automated processes can easily combat this type of piracy.[130] Companies like Audible Magic successfully employ an automated process to fight piracy when a website hosts links to complete files of copyright-protected music.[131] Audible Magic relies on automated content recognition, which automatically scans copyright-protected reference materials to create a "digital fingerprint" for that audio file.[132] This technology essentially records how the file would sound to the human ear, and then uses this reference file to detect matches that vary in "file formats, codecs, bitrates, and compression techniques."[133] Using this automated process, Audible Magic is able to accurately find pirated audio clips in seconds.[134] Without

---

125. Press Release, *supra* note 120.

126. Janet Novak, *How SEC's New Robocop Profiles Companies for Accounting Fraud*, FORBES (Aug. 9, 2013), http://www.forbes.com/sites/janetnovack/2013/08/09/how-secs-new-robocop-profiles-companies-for-accounting-fraud/, *archived at* http://perma.cc/HHD8-7WL8.

127. Jacobellis v. State of Ohio, 378 U.S. 184, 188 (1964).

128. *Id*. at 197.

129. *See, e.g.,* Joe Marine, *Vimeo Uploads with Copyright Material Will Now be Flagged Automatically, Even if It's Licensed*, NOFILMSCHOOL.COM (May 22, 2014), http://nofilmschool.com/2014/05/upload-vimeo-copyright-match-audible-magic, *archived at* http://perma.cc/274U-AFF4.

130. *See, e.g.*, *Media Identification*, AUDIBLE MAGIC, http://www.audiblemagiccorp.com/media-identification/ (last visited Oct. 9, 2014), *archived at* http://perma.cc/FW6C-JXUC.

131. *Id.*

132. *Content Registration*, AUDIBLE MAGIC, http://www.audiblemagiccorp.com/content-registration/ (last visited Oct. 9, 2014), *archived at* http://perma.cc/P8MY-3F8P.

133. *Media Identification*, *supra* note 130.

134. *Broadcast Monitoring*, AUDIBLE MAGIC, https://www.audiblemagic.com/broadcast-

question, Audible Magic's automated system, and others like it, have helped fight piracy. Discounting the value of these automated programs because of several high-profile blunders would be a mistake because there are many occasions in which piracy is unambiguous. When a movie or song is copied in its entirety, an automated process is a valuable tool.

However, not all cases of piracy involve unadulterated copying of source material.[135] When copyrighted material is used as a part of a larger artistic expression, the use is less like running a red light and more like distinguishing pornography from art.[136] In these situations there are few clear rules, and automation is less accurate.[137]

Fair use is the best example of where automation consistently fails. Fair use is difficult to clearly define, and the Second Circuit has declared, "the issue of fair use . . . is the most troublesome in the whole law of copyright."[138] The four factors a court will use when deciding whether a use is fair are frequently described as subjective and difficult to apply uniformly.[139] The United States Supreme Court has recognized that, while the factors have been clearly stated, the task of identifying fair use "is not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis."[140] The Seventh Circuit has gone so far to explicitly state that the factors in the fair use doctrine "do not constitute an algorithm that enables decisions to be ground out mechanically."[141]

In a study of the fair use doctrine, Georgetown Law Professor Dan L. Burk and University of Minnesota Professor of Law, Julie E. Cohen, concluded that a fully-automated process will never be able to fully account for fair use, stating, "[i]n reality, an algorithm-based approach to fair use is unlikely to accommodate even the shadow of fair use as formulated in current copyright law."[142] They deduced that creating a program that attempted to account for fair use would "require both a bewildering degree of complexity and an impossible level of prescience."[143] They further concluded that even if such a program was devised, it would quickly become prone to error because "fair use is a dynamic, equitable doctrine designed to respond to changing conditions of use."[144]

Fair use is too fluid and amorphous to ever be truly captured by a computer

---

monitoring/ (last visited Nov. 27, 2014), *archived at* http://perma.cc/4XNM-6XX7.

135.   Burk & Cohen, *supra* note 109, at 55-56.

136.   *Id.*

137.   *Id.*

138.   Dellar v. Samuel Goldwyn, Inc., 104 F.2d 661, 662 (2d Cir. 1939).

139.   *See* Ned Snow, *Fair Use as a Matter of Law,* 89 DEN. U. L. REV. 1, 28-29 (2011) (stating,"'fairness' is an inherently subjective concept").

140.   Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994).

141.   Chicago Bd. of Educ. v. Substance, Inc., 354 F.3d 624, 629 (7th Cir. 2003).

142.   Burk & Cohen, *supra* note 109, at 55.

143.   *Id*. at 56.

144.   *Id.*

program.[145] It requires an analysis of subjective factors, such as whether the use was "transformative."[146] Therefore, businesses that seek to fight piracy must acknowledge that implementation of a fully-automated process would likely catch blatant cases of copyright infringement but would also likely trample on cases of fair use.

A. Liability for Filing False DMCA Takedown Requests Under Section 512(f)

If a business implements an automated system that does not adequately account for fair use, what legal liabilities may be incurred? The DMCA includes section 512(f), which states that any party who "knowingly materially misrepresents" in the DMCA takedown request is "liable for any damages, including costs and attorneys' fees incurred by the alleged infringer."[147] Thus, copyright-holders may be liable if they file DMCA takedown requests when they know the particular use does not violate any of their rights as a copyright holder.[148] The material misrepresentation standard has been interpreted to "not impose liability for issuing a defective notice *per se,* only for making false claims of infringement."[149] Sending an incomplete takedown notice, or submitting one that does not properly identify the infringing material does not reach the standard of misrepresentation required by section 512(f).[150] Additionally, plaintiffs must establish that they suffered material harm from a misrepresented DMCA takedown request.[151] In fact, the District Court of Maryland willingly conceded that a copyright holder may have knowingly misrepresented his rights on a DMCA takedown request, but after the plaintiff could not prove any damages, the claim was dismissed.[152]

Despite the widespread use of DMCA takedown requests, litigation under section 512(f) for false requests rarely occurs.[153] There have been only a handful of published cases in which a section 512(f) claim is at issue—out of the hundreds of millions of URLS that have been removed due to DMCA requests.[154] While addressing the issue of liability under section 512(f), the District Court of Maryland concluded, "There is not a great deal of case law interpreting this provision . . . ."[155] Because liability under section 512(f) has been rarely litigated,

---

145. *Id.* at 55.

146. 17 U.S.C. § 107 (2013).

147. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512(c)(3) (2014).

148. *Id.*

149. Disney Enters., Inc. v. Hotfile Corp., No. 11-20427-CIV., 2013 WL 6336286, at *46 (S.D. Fla. Sept. 20, 2013).

150. Arista Records, Inc. v. Mp3Board, Inc., No. 00 CIV. 4660 (SHS), 2002 WL 1997918, at *15 (S.D.N.Y. Aug. 29, 2002).

151. Ground Zero Museum Workshop v. Wilson, 813 F. Supp. 2d 678, 704 (D. Md. 2011).

152. *Id.* at 710.

153. *Id.*

154. *Id.*

155. *Id.*

the legal standards for section 512(f) are still being established. The following analysis is intended to provide a brief overview of what kind of liability a business may incur if it employs an automated system for generating DMCA takedown requests.

 1. *Knowledge Requirement for Section 512(f) Liability.*—Although section 512(f) is intended to punish parties who "knowingly" misrepresent their rights in a DMCA takedown request, courts are still settling what kind of knowledge is required for liability under section 512(f).[156] Does this law require actual knowledge by copyright holders that the particular use does not violate their rights, or does it just require that they reasonably should have known that the use does not violate their copyright?

 The Northern District of California addressed this issue when two young men sued Diebold Incorporated claiming violation of section 512(f) after Diebold sought to remove a posting of several internal company emails, which these men posted online.[157] The court stated that Diebold actually knew, or "should have known if it acted with reasonable care or diligence" that this material did not infringe on any copyrights.[158] Under this standard of constructive knowledge, Diebold was found liable under section 512(f).[159] This case established a clear precedent that a copyright holder may be liable for misrepresentation without actual knowledge that their DMCA takedown request asked for the removal of non-infringing material.[160] Rather than require actual knowledge for liability, the court could find liability if the party sending the request "should have known" that such a request was an error.[161]

 Shortly after this decision was made, the Ninth Circuit disregarded this standard when it found the Motion Picture Association of America ("MPAA") not liable for a section 512(f) suit after they filed a DMCA takedown against a website that claimed to host downloads of full-length pirated movies; however, the website did not actually host any such pirated material.[162] The Ninth Circuit held that section 512(f) requires *actual* knowledge of misrepresentation, stating that "[a] copyright owner cannot be liable simply because an unknowing mistake is made, even if the copyright owner acted unreasonably in making the mistake . . . . Rather, there must be a demonstration of some actual knowledge of misrepresentation on the part of the copyright owner."[163] The court held that the MPAA had a subjective, good faith belief that there was infringement and was therefore not liable under section 512(f).[164]

 Several subsequent cases outside the Ninth Circuit applied this subjective,

---

156. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512(c)(3) (2014).
157. Online Policy Group v. Diebold Inc., 337 F. Supp. 2d 1195, 1198 (N.D. Cal. 2004).
158. *Id.* at 1204.
159. *Id.*
160. *Id.*
161. *Id.*
162. Rossi v. Motion Picture Assoc. of Am. Inc., 391 F.3d 1000, 1006 (9th Cir. 2004).
163. *Id.* at 1005 (citation omitted).
164. *Id.*

good faith belief standard that requires actual knowledge of misrepresentation. In *Dudnikov v. MGA Entertainment, Inc.*, a husband and wife claimed MGA violated section 512(f) after MGA terminated their eBay listing claiming copyright infringement.[165] The United States District Court for the District of Colorado relied on the *Rossi* precedent and held that the plaintiffs' claim required "substantial evidence that MGA knowingly and materially misrepresented" the purported infringement by the plaintiffs.[166] The court ultimately held that MGA was not liable under section 512(f).[167] Again, in *Third Education Group, Inc. v. Phelps*, a federal district court in Wisconsin was asked to consider the standard that should be applied after a partnership dissolved and one partner copied extensively from the material of the former partnership's website.[168] Citing *Rossi,* the court held a party cannot be found liable for misrepresentation under the DMCA "because of an unknowing mistake, even if that mistake [is] objectively unreasonable."[169] In *Cabell v. Zimmerman*, the United States District Court for the Southern District of New York also required actual knowledge after a man filed a section 512(f) claim after the Actor's Equity Association removed his video from YouTube through a DMCA takedown request.[170] The court stated that a prerequisite for section 512(f) liability is that "a defendant must have actual knowledge that it is making a misrepresentation of fact."[171] Although few courts have had a chance to decide whether section 512(f) requires actual knowledge, the Ninth Circuit seems to have set the standard. Future courts will likely follow suit and require actual knowledge.[172]

These cases have only addressed occasions in which a *person* sends a DMCA request.[173] Under those circumstances it is likely that the party filing a DMCA takedown request will be liable only if the request is sent with actual knowledge that the information is not infringing. In *Ouellette v. Viacom International, Inc.*, the United States District Court for the District of Montana specifically addressed how the "actual knowledge" requirement under section 512(f) applies when a copyright holder employs an automated system for generating takedown

---

165. Dudniko v. MGA Entertainment, Inc., 410 F. Supp. 2d 1010, 1011 (D. Colo. 2005).

166. *Id.* at 1012.

167. *Id.* at 1013.

168. Third Educ. Group, Inc. v. Phelps, 675 F. Supp. 2d 916, 920 (E.D. Wis. 2009).

169. *Id.* at 927.

170. Cabell v. Zimmerman, No. 09 Civ. 10134(CM), 2010 WL 996007 at *4 (S.D.N.Y. Mar. 12, 2010).

171. *Id.* (emphasis omitted).

172. *See* Disney Enter., Inc. v. Hotfile Corp., No. 11-20427-CIV., 2013 WL 6336286 at *46 (S.D. Fla. Sept. 20, 2013) (both parties conceded that actual, subjective knowledge of misrepresentation was required for section 512(f) liability "based upon the theory that one cannot knowingly misrepresent what one does not understand to be false.").

173. In *Rossi* the MPAA admitted that it "relied heavily" on an automated program to locate piracy, but the website that was a party in that case was found with human review. Thus, the court did not analyze the merits of MPAA's automated program. Rossi v. Motion Picture Assoc. of Am. Inc., 391 F.3d 1000, 1005 n.7 (9th Cir. 2004).

requests.[174] In *Ouellette*, a young woman filed a section 512(f) claim against Viacom for removing a video she posted on YouTube, claiming Viacom misrepresented itself in its DMCA takedown request.[175] The court stated that the plaintiff's pleading under *Rossi* required a demonstration that Viacom actually knew that the plaintiff's videos were fair use but nevertheless issued the takedown notice.[176] The plaintiff, however, argued that Viacom used "scanning software" that has a history of abusing the takedown process.[177] While this court did not object to the finding that Viacom's scanning software may have been prone to error, the judge concluded that under *Rossi* the plaintiff was still "required to plead facts that Viacom knew that its 'scanning software' was flagging her non-infringing videos and that Viacom issued a takedown notice nonetheless."[178]

For now, it appears that *Rossi* is setting the standard that is likely to be applied to the "knowledge" requirement under section 512(f). While it is yet to be settled by many courts, the majority of courts that have addressed the issue of knowledge under section 512(f) have required actual subjective knowledge of wrongdoing.[179] By setting the standard for "knowledge" under section 512(f) so high, courts have essentially incentivized copyright holders *not* to gain any knowledge about the DMCA request being generated. This approach raises its own liability issues.

2. Liability for Failing to Account for Fair Use Through Automation.—Since liability under section 512(f) requires actual subjective knowledge, businesses may be tempted to avoid accounting for fair use at all, in order to remain ignorant of any abuses that may occur. Can a corporation be found liable under section 512(f) after employing an automated program that simply does not take into account fair use at all? This question has become a corollary issue to the actual knowledge requirement in several court cases. More specifically, several plaintiffs have argued that the sender of a DMCA takedown request has avoided gaining actual knowledge of misrepresentation by simply never attempting to analyze takedown requests for fair use at all.[180]

The few courts that have addressed this murky question have provided very few clear rules. In *Lenz v. Universal Music Corp.*, a woman brought section 512(f) claims against Universal Music Corporation after the company filed a

---

174.   Ouellett v. Viacom Intern., Inc., No. CV 10-133-M-DWM-JCL, 2012 WL 850921, at *3 (D. Mont. Apr. 25, 2012).

175.   *Id.* at *4.

176.   *Id.*

177.   *Id.*

178.   *Id.*

179.   *See generally* Ross Drath, *Hotfile, Megaupload, and the Future of Copyright on the Internet: What Can Cyberlockers Tell Us about DMCA Reform?*, 12 J. Marshall Rev. Intell. Prop. L. 205 (2012) (discussing the knowledge requirement concerning liability for contributory negligence in hosting infringing material).

180.   *See, e.g.*, Lenz v. Universal Music Corp., 572 F. Supp. 2d 1150, 1151-52 (N.D. Cal. 2008).

DMCA takedown request with YouTube to remove a twenty-nine second video she posted of her son dancing to a song by Prince without Universal first considering whether this was fair use.[181] When Universal moved to dismiss the case, the Northern District of California court denied the motion saying that a good faith DMCA takedown request requires that "the owner must evaluate whether the material makes fair use of the copyright."[182] The court's language left little doubt that a party must verify whether a use of copyright-protected material was in fact fair use before the copyright holder issues a DMCA takedown request.[183] Yet five years after the first iteration made in *Lenz*, the court substantially retreated from this position, stating "mere failure to consider fair use would be insufficient to give rise to liability under [section] 512(f)."[184] Rather, the court returned to *Rossi* and held that Universal would need "actual knowledge" that the takedown contained a misrepresentation.[185] Several months after the *Lenz* opinion was issued, the District Court of Massachusetts also supported the standard of not requiring copyright holders to account for fair use before filing a DMCA takedown request.[186] The court held that a copyright holder did not have to consider fair use before filing a DMCA request because doing so would be "at odds with Congress's express intent of creating an 'expeditious,' 'rapid response' to 'potential infringement' on the Internet."[187] The court conceded that not requiring an examination of fair use may hurt a party's ability to make fair use of another's property, but retorted that "resetting the balance is for Congress and not the court to strike."[188] Ultimately, the court concluded that it was not necessary to account for fair use before sending a DMCA takedown request.[189]

The most instructive case on the issue of willful blindness of fair use, especially in the context of automation, was made by the United States District Court for the Southern District of Florida in September 2013.[190] In *Disney Enterprises, Inc. v. Hotfile Corp.*, Warner Brothers ("Warner") sent Hotfile, an offshore file storage company, over 400,000 DMCA takedown requests.[191] Warner admitted that at least 600 of these takedown requests were in error.[192] As

---

181. *Id.*

182. *Id.* at 1154.

183. *Id.*

184. Lenz v. Universal Music Corp., No. 5:07-cv-03783-JF, 2013 WL 271673, at *6 (N.D. Cal. Jan. 24, 2013) (emphasis added).

185. *Id.*

186. Tuteur v. Crosley-Cocoran, 961 F. Supp. 2d. 333, 343-44 (D. Mass. 2013).

187. *Id.* at 344.

188. *Id.*

189. *Id.*

190. Disney Enter., Inc. v. Hotfile Corp., No. 11-20427-CIV., 2013 WL 6336286 (S.D. Fla. Sept. 20, 2013).

191. *Id.* at *15.

192. *Id.*

a result, Hotfile filed a counterclaim against Warner under section 512(f).[193]

In its counterclaim, Hotfile argued that Warner employed an automated process for generating DMCA takedown requests that "prevented it from *acquiring* subjective knowledge" of misleading DMCA takedown requests.[194] The process Warner used to generate these DMCA requests is similar to the approach used by many businesses discussed throughout this Note. First, Warner hired a third-party to manage "anti-piracy enforcement," and that entity used an "automated review process."[195] Warner also hired employees to verify takedown requests, but these employees only verified that the site was used for Internet piracy and that the link contained content that infringed Warner's copyright.[196] The court pointed out that during the human review stage, "Warner was concerned with determining whether it owned the works rather than whether the use of the works infringed on its copyrights to support a proper section 512(c) claim."[197] The system Warner used seemed to rely heavily on an automated search process in which humans tailored the search terms and conducted "spot checks" on the requests that were generated.[198]

It is this largely automated system that led to at least 600 DMCA takedown requests sent in error.[199] In fact, the court acknowledged that "Warner readily admits that mistakes do occur."[200] Warner's employees never downloaded or even reviewed "any Hotfile content before marking it for removal."[201] Hotfile cited an occasion in which Warner sent a DMCA takedown request on behalf of content to which it did not even own the copyrights.[202] Hotfile cited more examples demonstrating that Warner's anti-piracy procedures were prone to error, but most of these examples were redacted out of the published opinion.[203] Although Warner's procedure had known flaws that were "readily" admitted, Warner repeatedly asserted that "its methodology and system features are common in its industry."[204] This assertion is likely true; thus the court's analysis provides valuable insight as to the standard that will be applied to other large

---

193. *Id.* at *46.

194. *Id.* at *47.

195. *Id.* at *14.

196. *Id.*

197. *Id.* at *47.

198. *Id.* at *14.

199. *Id.* at *15.

200. *Id.*

201. *Id.*

202. *Id.* at *16.

203. *See id.* at *15 (for example, one paragraph in its entirety reads: "Hotfile also points to instances of anecdotal errors to show how unsound Warner's search practices might have been. For example, [Redacted] (Thompson Decl. Ex. 4 (Kaplan Dep.) at 16:10-17:4 (DE 354-5).) Warner also apparently [Redacted] (Thompson Decl. Ex. 5 (DE 304-1 (filed under seal); DE 354-6).) A search for the [Redacted] (Thompson Decl. Ex. 16 (DE 304-6 (filed under seal); DE 354-17).) And [Redacted] (Thompson Decl. Ex. 18 (DE 304-6 (filed under seal); DE 354-19).)").

204. *Id.* at *14-15.

businesses that use automation in generating DMCA takedown requests.

In its decision regarding Hotfile's section 512(f) claim, the court concluded that the claim should survive summary judgment and be presented to the jury.[205] The court addressed the issue of using automation to avoid actual knowledge of wrongdoing stating, "Warner's reliance on technology to accomplish the task might prevent it from forming any belief at all."[206] The court went on to acknowledge that this issue would be one of first impression, stating that the court was "unaware of any decision to date that actually addressed the need for human review, and the statute does not specify how belief of infringement may be formed or what knowledge may be chargeable to the notifying entity."[207] The court, however, did not make a definitive ruling about whether this violates the nature of section 512(f), stating what while the issues of willful blindness through automation and the liability that this process creates are "engaging questions," there was sufficient evidence that Warner "intentionally targeted files it knew it had no right to remove."[208] Despite not making a definitive ruling on the issue of willful blindness through automation, the court's analysis demonstrates that a business cannot expect a de facto pardon from the requirement of actual knowledge if it instead chooses to implement an automated process with known flaws. On the contrary, this court's decision demonstrates that a court will look into a DMCA takedown review process and may ultimately leave the determination of section 512(f) liability for a jury.

B. Avoiding Liability Under Section 512(f)

These cases provide a legal framework for businesses seeking a policy for handling DMCA takedowns that offers the least liability. First, any procedure used to respond to DMCA takedown requests must respond to the requests quickly in order for that party to remain within the safe harbor provisions of the DMCA.[209] For this reason, automation, as least to some degree, seems to be required. Through automation, copyright owners do not have to personally scour the internet in an endless game of "Whac-A-Mole"[210] searching for instances of piracy. Automation also provides ISPs the ability to respond to these requests quickly in order to keep up with the onslaught of requests that many ISPs face on a regular basis.[211] Under the first requirement, automation seems like an ideal solution.

In addition, the ideal procedure would attempt to account for fair use. A

---

205. *Id.* at *48.

206. *Id.* at *47.

207. *Id.*

208. *Id.* at *48.

209. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512(c)(1)(C) (2010).

210. Nate Anderson, *Rightsholders Tire of Takedown Whac-A-Mole, Seek Gov't Help*, ARS TECHNICA (May 4, 2010, 9:05 AM), http://arstechnica.com/tech-policy/2010/05/rightsholders-tire-of-takedown-whac-a-mole-seek-govt-help/, *archived at* http://perma.cc/6NMD-VBXY.

211. Cushing, *supra* note 23.

procedure that ignored possible occasions of fair use through automation would stray into dangerous and uncharted legal territories. An autonomous procedure that avoids fair use considerations would likely be heavily scrutinized by the court, and may ultimately be left for the jury to decide whether there will be liability under section 512(f).[212] Under this secondary consideration, automation would likely lead to increased litigation and possibly increased liability under section 512(f) and is thus, not an ideal solution.

These parameters define the precarious legal tightrope that all DMCA procedures must walk. Rely on automation too heavily, and fair use is likely to be trampled and liability under section 512(f) increases. Avoid automation altogether, and response time significantly decreases and the business could allow piracy to continue unabated or may even find itself outside of safe harbor protections and liable for hosting pirated material. Leaning too far one way or the other always increases liability. The ideal solution balances these competing needs in a solution that utilizes automation, but also protects fair use.

IV. FINDING A SOLUTION THROUGH CROWDSOURCING

"Crowdsourcing" is defined as "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers."[213] Large ISPs that receive thousands of DMCA requests, yet hire employees or utilize computer programs to filter out copyright-protected material, are ignoring their greatest asset—users that created the content. Every minute users are uploading over 100 hours of content to YouTube.[214] These videos are then viewed at a rate of over six billion hours per month.[215] Furthermore, many viewers not only watch these videos, but also provide feedback on the video. More than half of all the videos on YouTube are either rated or commented upon by the users.[216]

These are billions of users that could be utilized to shoulder the enormous task of separating piracy and fair use. Rather than employing a small cadre of employees hopelessly looking for piracy, or utilizing a fully automated system that will certainly flag non-infringing material, an ISP can ask its own users to solve this problem. Why have one person watching millions of videos for piracy when you can have millions of people watching one video? By placing the burden of identifying the material on the users, ISPs are able to make the task far

---

212. *See* Disney Enter., Inc. v. Hotfile Corp., No. 11-20427-CIV., 2013 WL 6336286 (S.D. Fla. Sept. 20, 2013).

213. *Crowdsourcing,* MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/crowdsourcing (last visited Oct. 12, 2014), *archived at* http://perma.cc/2G22-9K2V.

214. *Statistics*, *supra* note 90.

215. *Id.*

216. *Latest YouTube Facts and Statistics*, TECHWELKIN (Mar. 3, 2012), http://techwelkin.com/latest-youtube-fact-and-statistics-infographic#.UkSl6oasim4, *archived at* http://perma.cc/84R7-KJHQ.

more manageable.

A. How Would Crowdsourcing Work?

First, ISPs would devise programs like YouTube's Content ID that would separate clearly infringing works from works that present less clear cases of piracy. This Note recommends following the Electronic Frontier Foundation's three strike policy[217] for filtering the material initially. The Electronic Frontier Foundation ("EFF") is a nonprofit legal organization that advocates for free speech, especially in areas of expanding technology.[218] The EFF advocates for a policy in which material would be removed when more than ninety percent of that material matches the audio track and video track of copyright-protected material.[219] If material is automatically filtered out, human creators should be given the opportunity to dispute this finding.[220] This system would eliminate the instances of clearly infringing material, the red-light-running material, quickly and easily.

Using this fully automated system, user-generated content could be separated into three categories: (1) material with no violations detected; (2) material with many violations detected; and (3) material with some violations detected. Under this scheme, material that does not contain any copyrighted material would be hosted without any interference. Other material that contains many violations and is blatant piracy will be removed automatically following the procedure advocated for by the EFF.[221] It is only the third category of material, content that contains some copyright-protected material, but is less than a ninety-percent match of audio and video, that would receive further consideration.[222]

Content that falls under the third category would be hosted and accessible, but would have a survey at the end of the material that asks for user feedback as to whether the material is fair use or a copyright infringement. For example, after the first fifty viewers of an online video watch it, the segment of the screen that was devoted to the video could then be replaced with a survey that notifies the user that it must be completed if the video is going to remain accessible. If the content is a printed article, the website could install a pop-up or a survey that becomes available when the user scrolls to the bottom of the article. This survey would take less than fifteen seconds to complete and would only be required of the first fifty consumers of the content. This option would be kept available for future consumers, but may not be as prominent. Once a website has the data from

---

217. *Fair Use Principles for User Generated Video Content*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/pages/fair-use-principles-user-generated-video-content (last visited Oct. 12, 2014), *archived at* http://perma.cc/HY25-PDL2.

218. *About EFF*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/about (last visited Feb. 21, 2014), *archived at* http://perma.cc/RSZ4-WZDR.

219. *Fair Use Principles for User Generated Video Content*, *supra* note 217.

220. *Id.*

221. *Id.*

222. *Id.*

at least fifty, or even one thousand users, the party hosting the content could better determine whether something is fair use without using any additional resources.

B. Using Empirical Data to Predict Fair Use?

Theoretically, the fair use doctrine is complex. The doctrine has few bright-line rules and is hard to precisely define. Empirical studies of the fair use doctrine, however, have revealed a completely different picture of fair use. These studies have detected consistent trends and patterns in the application of fair use. Using empirical evidence, the fair use doctrine can be not only understood, but also predicted. ISPs do not need to determine whether the material *should* be considered fair use, rather, they only need to determine whether *a judge is likely to consider* the material fair use. This is a matter of simple probabilities.

Professor Barton Beebe, in *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, studied every published opinion that considered fair use from 1978-2005 and found that judges regularly fall into patterns with their decisions and follow predictable trends when deciding fair use cases.[223] Building on his work, Pamela Samuelson concluded, after her own empirical study of the data, that "copyright fair use case law is more coherent and more predictable than many commentators seem to believe. Fair use cases tend to fall into common patterns . . . ."[224] Lastly, Matthew Sag's empirical study of fair use, the most recent study available, agreed with this conclusion, saying, "[i]s fair use predictable? The evidence presented here . . . makes it difficult to sustain the common charge of incoherence and unpredictability."[225] Using human-provided information as well as data collected in these empirical studies, a computer program should be able to reliably predict the likelihood that something will be considered fair use within certain percentages. Using these percentages, the ISP can then decide whether to remove the material or keep it accessible.

The key to unlocking the puzzle of fair use is asking the human users to identify the character of the material through carefully crafted questions. First, the user would be asked to identify the purpose of the material. The user would be asked to "check all that apply" and would be provided with the following categories: "commercial purpose, criticizing the original creator, parody, biography, history, social/political criticism, use in court case, or a purpose not listed here."[226] These categories, except commercial use, were chosen because Neil Weinstock Netanel found that all of these categories "were held to be fair use to a statistically significant degree" in his review of the empirical studies regarding fair use.[227] In other words, if the purpose of the material was identified

---

223. Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549 (2008).

224. Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537 (2009).

225. Matthew Sag, *Predicting Fair Use*, 73 OHIO ST. L.J. 47, 79 (2012).

226. *See, e.g.*, Netanel, *supra* note 104, at 750.

227. *Id.*

to be one of these categories, there is a strong likelihood it would be considered fair use. These categories also exclude uses of copyright-protected material that have traditionally been considered fair use such as satire and news reporting, but have not generated statistically significant results for fair use.[228]

This is an essential step because it asks a human user to do something a computer program cannot. While an automated program may be able to tell when a video matches a reference file, that program cannot detect the purpose of this reproduction, and fair use allows for reproduction, depending on the purpose. Accurately determining the purpose behind a video is impossible for a computer program, but it is very easy for humans to determine. Using these categories, a computer program can then generate the likelihood that the use is fair use.

For example, Samuelson has found that, despite the United States Supreme Court's insistence in *Campbell v. Acuff-Rose Music Inc.* that parodies are not always fair use and must be considered on a case-by-case basis,[229] every time this issue has been adjudicated since *Campbell* every court has found parody to be fair use.[230] This computer program would therefore, create a strong correlation between cases in which many users identified the material to be a parody and the likelihood that the use was fair use. Using copyrighted material for commercial purposes, on the other hand, has been found to have a negative correlation with fair use. Beebe found that if a defendant has made a non-transformative commercial use of copyrighted material, that defendant only has a 35.5% chance of winning under the fair use defense.[231] Using this human-provided data, a program can then create probabilities to calculate the likelihood, to a precise percentage, that something is fair use.

This data can then be corroborated by the second stage of the survey. The second stage will ask a yes or no question: "Is this a transformative work that adds something new, with a further purpose or different character, altering the original with new expression, meaning, or message?" This will be followed by a question that asks, "After viewing this material how likely are you to view the original?" and allows the user to choose between "more likely," "less likely," or "no change." This question corresponds with the fourth factor of the fair use factor test and is highly correlative with a finding of fair use.[232] Professor Beebe found that the outcome of this factor "coincided with the outcome of the overall test in 83.8% of the 297 dispositive opinions."[233] Sag came to a similar conclusion about the power of finding a transformative use, stating that when a transformative use is present, the chances of finding fair use doubles.[234] In fact, he found that "not knowing anything else about the defendant's use, a plaintiff can expect to win a clear majority of cases where there is no indication of

---

228. *Id.*

229. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994).

230. Samuelson, *supra* note 224, at 2550.

231. Beebe, *supra* note 223, at 606.

232. 17 U.S.C. § 107 (2013).

233. Beebe, *supra* note 223, at 584.

234. Sag, *supra* note 225, at 76.

transformative use, but otherwise expect to lose all but 38% of the time."[235] In his review of the material, Netanel found that this correlation has even increased recently, stating, "I also found a consistently high rate of defendant wins, reaching 100% in 2001-2005 and 2006-2010, in those cases in which the court found that the defendant's use was, in fact, unequivocally transformative."[236] While identifying the use as "transformative" is simply impossible for a computer program, it can be accomplished quite easily by a human who has just consumed the material. The human user simply has to identify whether the work meets the definition of "transformative" supplied by the United States Supreme Court.[237] If this definition is unclear to many users, an ISP could explain it further based on the many corollary definitions supplied in case law. If the survey finds that many users consistently identify a work as transformative, there is a probability nearing one hundred percent that this work is fair use. This conclusion is further corroborated by the second question, which essentially asks the user to identify the effect this use will have on the market for the original. Determining the harm this use will have on the original is one of the key balancing decisions most judges make in determining fair use.[238] If the second question reveals that users are more likely to view the original after viewing this material, judges are far less likely to find that this use hurts the original and is therefore likely to be fair use.[239]

Completing these questions would require no typing on the part of the human user and could be accomplished with three clicks of the mouse in less than fifteen seconds. However, this critical human input is the essential ingredient for an automated program to truly account for fair use because it is the key step a computer program cannot make. Once a human provides a program with this data, that program is able to calculate based on empirical data the chances that the material represents fair use to a percentage point. An ISP can then choose to remove the material or leave it accessible based on this percentage. ISPs will then be able to actively combat piracy while protecting legitimate use with a high degree of accuracy.

C. Is Crowdsourcing Effective?

Can people really be trusted to honestly identify piracy?   Under this approach, it is possible that people who understand the system would intentionally lie in their responses in order to keep the material available. For instance, if someone watched a pirated episode of *Seinfeld*, they could then identify the source as a parody and increase the chances that it would be deemed fair use. While such actions are always possible, it is possible to mitigate the

---

235. *Id.*

236. Netanel, *supra* note 104, at 754.

237. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (stating that a work is "transformative" if it "adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message").

238. Beebe, *supra* note 223, at 621.

239. *Id.*

negative consequence of these actions.   First, the automated process will automatically remove clearly infringing material. An episode of *Seinfeld* that is pirated in its entirety would likely be removed automatically.   Second, the computer program could maintain a database that stores the user's username or IP addresses and the feedback they have provided in the past.   If a user has consistently identified something as fair use when many others have identified it as piracy, that user's survey results can have a diminished value or perhaps no value at all. On the other hand, users who have consistently identified piracy or fair use with a high degree of accuracy can have their survey results given greater weight than the average user.   In this way, the survey results could take into account competency and weigh results accordingly. Lastly, if an ISP finds that some users are skewing the results, the ISP could simply extend the required survey to a greater audience.

The evidence of past crowdsourced projects has proven that crowdsourcing is able to deliver reliable results. Wikipedia, for example, was created with the noble purpose of allowing access to an encyclopedia to anyone with an Internet connection. Today, Wikipedia has had over one billion edits from more than twenty-seven million users.[240]   Users can access articles on quantum electrodynamics,[241] Merovingian art and architecture,[242] as well as a listing of all the minor characters to ever appear in Charles M. Schultz's *Peanuts* comic strip, complete with biographies.[243] More importantly, as early as 2005 Wikipedia was found to be about as accurate as the *Encyclopedia Britannica*.[244]

Crowdsourcing has continued to be adopted not only for simple tasks like reviewing restaurants or movies, but also has recently been the source of labor for many complex and vexing problems. Ordinary citizens have become powerful tools through crowdsourcing and are now identifying billions of galaxies,[245]

---

240.   John Cabell, *Ten Impressive, Weird and Amazing Facts About Wikipedia*, WIRED (Jan. 12, 2011, 1:29 PM), http://www.wired.com/business/2011/01/ten-facts-about-wikipedia/2/, *archived at* http://perma.cc/SH2X-X97B.

241.   *Quantum Electrodynamics*, WIKIPEDIA, http://en.wikipedia.org/wiki/Quantum_electrodynamics (last visited Oct. 12, 2014), *archived at* http://perma.cc/8MG8-K7SP.

242.   *Merovingian Art and Architecture*, WIKIPEDIA, http://en.wikipedia.org/wiki/Merovingian_art (last visited Oct. 12, 2014), *archived at* http://perma.cc/8MG8-K7SP.

243.   *List of Minor Characters in Peanuts*, WIKIPEDIA, http://en.wikipedia.org/wiki/List_of_minor_characters_in_Peanuts (last visited Oct. 12, 2014), *archived at* http://perma.cc/8X4B-SJ6M (a personal favorite character is "shut up and leave me alone"—a nameless, faceless kid who was Charlie's bunkmate during summer camp. The only thing he said throughout several comics was "shut up and leave me alone.").

244.   *Wikipedia Survives Research Test*, BBC NEWS (Dec. 15, 2005), http://news.bbc.co.uk/2/hi/technology/4530930.stm, *archived at* http://perma.cc/9JPF-E487.

245.   Sara Gates, *Citizen Scientists Classify Galaxies in Massive Crowdsourced Project, But Can You?*, HUFFINGTON POST (Sept. 26, 2013, 8:31 AM), http://www.huffingtonpost.com/2013/09/26/citizen-scientists-classify-galaxies-galaxy-zoo-2_n_3990747.html, *archived at* http://perma.cc/RYK8-X9PN.

classifying historical documents from the national archives,[246] detecting fraud in tax returns,[247] and even grading students' homework.[248] If enough users provide feedback, crowdsourcing is reliable.

D. Solutions for Copyright Holders

Better screening on the part of ISPs through crowdsourcing only addresses a solution for parties who *receive* DMCA takedown requests. How can the parties that generate these requests better protect fair use? What should copyright holders like Microsoft, Warner Brothers, and others do to better honor fair use while still expeditiously removing offending material? These companies could still automate their generation of DMCA requests, however, they should allow greater discretion for ISPs that employ their own automated processes through techniques like crowdsourcing to protect fair use. Copyright holders could adjust the automated programs they use in order to allow ISPs that have their own in-house processes for detecting infringements to be excluded from receiving DMCA requests unless there is blatant infringement.

While initially it may appear that allowing these ISPs greater discretion and generating fewer DMCA requests would result in diminished demand for the original product, this is not likely to be the case. First, part of the fair use test considers whether the use diminishes demand for the original.[249] The more users identify a work as making them less likely to view the original, the less likely that work will be protected by the fair use doctrine.

Second, if the material is kept accessible because it has been identified as fair it may actually become an additional source of revenue. In May 2008, Sony music released Chris Brown's single "Forever" and it quickly made it to Billboard's Top Ten on the "Hot List" of most popular songs.[250] Over a year later, a couple danced to this song during their wedding and posted the video on YouTube.[251] After this video was posted, Sony received notification that a

---

246. Nathan Raab, *Crowdsourcing Technology Offers Organizations New Ways to Engage Public In History*, FORBES (Sept. 20, 2013, 7:36 AM), http://www.forbes.com/sites/nathanraab/ 2013/09/20/technology-offers-organizations-new-ways-to-engage-the-public-in-history/, *archived at* http://perma.cc/BC97-J9QP.

247. Don Mackenzie, *How Crowdsourcing Fraud Detection Can Help Reduce Tax Cheats*, DATA INFORMED (Sept. 23, 2013, 11:40 AM), http://data-informed.com/crowdsourcing-fraud-detection-can-help-reduce-tax-cheats/, *archived at* http://perma.cc/L6LN-F2D4.

248. *First Trial of Crowdsourced Grading for Computer Science Homework,* MIT TECHNOLOGY REVIEW (Sept. 4, 2013), http://www.technologyreview.com/view/519001/first-trial-of-crowdsourced-grading-for-computer-science-homework/, *archived at* http://perma.cc/54YZ-C5JF.

249. 17 U.S.C. § 107 (2012).

250. *The Hot 100*, BILLBOARD (May 10, 2008), https://web.archive.org/web/20130524082734/ http://www.billboard.com/charts/2008-05-10/the-hot-100, *archived at* http://perma.cc/U46E-MMPA.

251. Margaret Gould Stewart, *Margaret Gould Stewart: How YouTube Thinks about*

YouTube video matched the audio of the copyright they held for the song and Sony could opt to remove the video.[252] Rather than strictly enforcing their copyright and potentially blocking fair use of their material, Sony chose to allow the video to remain accessible.[253] The video has since gone on to get over eighty-three million views[254] and eighteen months after Chris Brown's song fell off the charts it was launched back to iTune's top ten and began earning Sony additional revenue.[255] This video was then parodied by NBC's hit show *The Office* in their season finale and once again brought additional attention to the song and additional revenue to Sony.[256] Allowing fair use of copyright-protected material not only prevents a copyright holder from possibly being liable under section 512(f) for filing misleading DMCA takedown requests, but also might help a copyright holder's bottom line. Fair use by its very nature, draws attention back to the original and creates additional streams of revenue at no cost. Employing a system that accounts for fair use and relying on this system creates a win-win scenario that will diminish liability while increasing revenue for copyright holders.

E. Walking the Tightrope

Using crowdsourcing, a business can then meet the criteria outlined earlier for an efficient DMCA filtering process. First, this process is "expeditious" as required by law.[257] It is unclear what the word "expeditiously" legally requires because it has never been litigated in the context of DMCA takedown requests. Many times the material is removed within minutes of getting such a request. Using crowdsourcing will require that this term be interpreted not to measure time, but rather how quickly material is removed in relation to the rate at which it is consumed. This would be a more accurate and helpful definition. For instance, a video might be posted on a Monday, yet not be watched by anyone until the following Sunday. This would mean waiting a week before anyone provides feedback as to its fair use. Initially, this would not seem expeditious at all. This approach, however, measures what really matters—how many people consume potentially infringing material. If a video is posted, yet no one watches it, no harm could possibly be done. However, if a video is posted and gets over 1,000 views within the first hour, then a crowdsourcing approach would identify

---

Copyright, TED (Feb. 2010), http://www.ted.com/talks/margaret_stewart_how_youtube_thinks_about_copyright.html, *archived at* http://perma.cc/6SU2-N9RX.

　　252. *Id.*

　　253. *Id.*

　　254. *JK Wedding Entrance Dance*, YOUTUBE, http://www.youtube.com/watch?v=4-94JhLEiN0 (last visited Oct. 12, 2014), *archived at* http://perma.cc/5EAF-87ST.

　　255. *Viral Wedding Video Gives Chris Brown's 'Forever' a Boost*, BILLBOARD (July 29, 2009), http://www.billboard.com/biz/articles/news/1267294/viral-wedding-video-gives-chris-browns-forever-a-boost, *archived at* http://perma.cc/MZ8C-FPKX.

　　256. Stewart, *supra* note 251.

　　257. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512(c)(1)(C) (2012).

that material quickly and would require that it be removed within a matter of minutes if it is genuine piracy. Although no court has yet taken this approach, it is a logical extension of the purpose behind the DMCA.

Second, this process is accurate and incorporates fair use. Using only fifteen seconds of a user's time, an automated program can substantially increase its level of accuracy. Highly public blunders would be diminished. Users would be more comfortable uploading their creative content. Consumers would be more attracted to the ISP knowing that it is a haven for creative and critical uses of copyrighted material. And both ISPs and copyright holders would have decreased legal liabilities.

CONCLUSION

The landscape of the Internet has changed drastically since 1998. Today the Internet is not only a source of international commerce, but also is a place where families upload their home movies, politicians grow their campaigns, and critics malign incompetence. The World Wide Web has become the world's largest theater where artists debut their newest creations.

Protecting copyrighted material in the modern world has become a daunting task. Copyright holders are forced to constantly comb the ever-changing landscape of the Internet looking for possible infringement. This means every article written, video uploaded, and song played must be reviewed for infringement. This task is being accomplished by automated processes rather than by humans. In addition, many internet-based businesses that host third party content can expect to receive hundreds, if not thousands of DMCA takedown requests every day. Once again this process is being automated in order to simply keep up with the requests.

While the DMCA certainly has its flaws and needs improvement, the larger problem is that the Internet contains billions of pieces of copyright protected material. Some of that material is pirated and some of it is not. How will the pirated material be removed without affecting the non-infringing material? Internet users are constantly writing, filming, editing, recording, or designing new material and making it accessible through the Internet. There must be a system for separating material that is genuine fair use from piracy. Humans cannot do it alone because it would be too time consuming and computers cannot do it alone because it requires a degree of analysis that is exclusively human in nature. If this problem is not addressed soon, it will be left solely to automated processes.

In the early years of the twenty-first century, many Internet-based businesses faced a similar problem as a result of growing automation. Websites that hosted email services, online polls, and blogs were being inundated with computer programs that were designed to act like human users, create accounts, and then generate information with a commercial or sinister purpose. Famously, in November 1999, one website released an online poll asking users which school had the best graduate program in computer science.[258] Several schools created

---

258. *What Is CAPTCHA,* GOOGLE, http://www.captcha.net/ (last visited July 22,

programs that were designed to mimic human users and voted for their school thousands of times.[259] Websites were being flooded by computer programs that were acting as humans. Rather than try to create a computer program that could separate human users from computers without any human input, the solution to this problem was crowdsourcing.[260] Now, whenever performing any mundane function on the internet, such as signing up for email or paying a bill, a human is asked to simply look at a distorted text and type what word they believe it looks like.[261] This process takes several seconds, but it prevents computer programs from acting like humans.[262] By dispersing the task to hundreds of millions of people, the once daunting task of separating computer programs from humans has been made manageable. This was a major problem facing the growth of the Internet, and though the solution was a minor inconvenience to millions of people, it has been effectively solved.

Today, the Internet is facing the daunting problem of distinguishing piracy from fair use. And the solution lies in the millions of people who consume the material on the Internet. Users must once again accept a minor inconvenience that is spread over millions of people in order to ensure that the Internet remains safe for both commerce and artistic expression. Through crowdsourcing, advanced computer programs can successfully discover if material is fair use with a high degree of accuracy based on human input. When millions of people answer several short questions in less than fifteen seconds, an advanced computer system can then take this information and extrapolate the likelihood that it is fair use based on empirical data. Most importantly, the users who are already consuming the material, rather than employees or computer programs, are identifying and categorizing the material.

When fair use is protected through a highly sophisticated and reliable process, ISPs will be less inundated with DMCA requests. Copyright holders will be less likely to be found liable under section 512(f) for filing misleading DMCA takedown requests, and the need to file these requests will be greatly diminished because a more reliable system has been created. Websites that adequately protect fair use will be more likely to receive new creative material from users. These users will be more likely to upload their material to these websites because they know their material will be protected. These websites can increase revenue by becoming havens for artistic expression. Furthermore, copyright holders who allow fair use of their copyright-protected material will more likely receive additional streams of revenue as users take the copyright-protected material and transform it. And most importantly, through crowdsourcing the consumers of the Internet, not computer programs, will determine the composition of the Internet. The future of the Internet would be limited not by the accuracy of our computer programs, but rather by the power of our imagination.

2014), *archived at* http://perma.cc/2YZ4-HVXM.

    259. *Id.*

    260. *Id.*

    261. *Id.*

    262. *Id.*