NUS Law Working Paper 2021/018

NUS Centre for Technology, Robotics, Artificial Intelligence & the Law Working Paper 21/01

Artificial Intelligence and Information Intermediaries

Daniel Seng

Associate Professor, Director, Centre for Technology, Robotics, Artificial Intelligence & the Law, Faculty of Law, NUS

[October 2021]

Artificial Intelligence and Information Intermediaries

Daniel Seng*

Abstract

The explosive growth of the Internet was supported by the Communications Decency Act (CDA) and the Digital Millennium Copyright Act (DMCA). Together, these pieces of legislation have been credited with shielding Internet intermediaries from onerous liabilities, and, in doing so, enabled the Internet to flourish. However, the use of machine learning systems by Internet intermediaries in their businesses threatens to upend this delicate legal balance. Will this affect the intermediaries' CDA and DMCA immunities, or expose them to greater liability for their actions? Drawing on both substantive and empirical research, this paper concludes that automation, as used by intermediaries, largely reinforces their immunities. In consequence, intermediaries are left with little incentive to exercise their discretion to filter out illicit, harmful and invalid content. These developments brought about by AI are worrisome and require a careful recalibration of the immunity rules in both the CDA and DMCA to ensure the continued relevance of these rules.

Introduction

1. Technology afficionados assert that the Internet of today was conceived with the commercialization of Internet infrastructure in 1995.[1] However, technology lawyers credit the enactment of the U.S. Communications Decency Act of 1996[2] and the Digital Millennium Copyright Act of 1998[3] for spurring the development of the commercial Internet. These two pieces of legislation, and their equivalents in the EU Electronic Commerce Directive[4], were attempts at providing answers to the pivotal question of whether an Internet intermediary – a company that

* Associate Professor, Faculty of Law, and Director, Centre for Technology, Robotics, AI and the Law (TRAIL), National University of Singapore. I would like to thank Ms Hitomi Yap and Mr Shaun Lim for their help with the research and editing of this paper. All errors and omissions however remain mine.

[1] Wikipedia, National Science Foundation Network, <https://en.wikipedia.org/wiki/National_Science_Foundation_Network#Privatization_and_a_new_network_architecture>.

[2] Title V, U.S. Telecommunications Act of 1996 (Pub.L. No. 104-104, 110 Stat. 56); codified as 47 U.S.C. § 230 (CDA).

[3] Pub. L. 105-304, 112 Stat. 2860 (1998); enacted as §§ 512, 1201-1205, 1301-1332 of Title 17 of the U.S. Code (DMCA).

[4] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market OJ L 178 (17 July 2000) (Directive on electronic commerce).

does not create or control the content of information – is liable for merely providing access to such information. In *Religious Technology Centre v. Netcom On-Line Communication Services Inc.*, the court provided the following answer in relation to claims against the intermediary for copyright infringement:

> No purpose would be served by holding liable those who have *no ability to control the information to which their [uploaders] have access*, even though they might be in some sense helping to achieve the Internet's *automatic* "public distribution" and the users' "public" display of files.… Where the infringing [uploader] is clearly directly liable for the same act, it does not make sense to adopt a rule that could lead to the liability of countless parties whose role in the infringement is nothing more than setting up and operating a system that is necessary for the functioning of the Internet. Such a result is unnecessary as there is already a party directly liable for causing the copies to be made.[5] (emphasis added)

2. Because the dissemination of content by the intermediary was considered automatic and caused by the uploader, the intermediary was held not liable.[6] This very same premise was applied to claims outside of copyright law. In an action against an intermediary for disseminating and publishing a defamatory statement, the court in Zeran v. America Online Inc. reached the same conclusion:

> The *amount of information communicated via interactive computer services is therefore staggering*. The specter of tort liability in an area of such prolific speech would have an obvious chilling effect. It would be *impossible for service providers to screen each of their millions of postings for possible problems*. Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted. Congress considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect.[7]

[5] 907 F.Supp. 1361 (N.D. Cal. 1995) (emphasis added).
[6] Ibid 1381-1382.
[7] 129 F.3d 327, 331 (4th Cir. 1997) (emphasis added).

3. This basic premise – that an intermediary is not liable for providing automated services to disseminate content which it was not involved in creating – has been codified in s 512 of the DMCA[8] and in s 230 of the CDA.[9] With these twin rules, Internet service providers and hosting companies are generally absolved of liability for disseminating illicit content such as copyright-infringing material, pornography, hate speech, and defamatory content authored by third parties.[10] These rules made it possible for websites, blogs, and social networks to host their users' content whilst being protected "against a range of laws that might otherwise hold them legally responsible for what their users say and do."[11] The rules have protected YouTube from copyright infringement for making available video clips shared by its users,[12] shielded Yelp from lawsuits for its users' negative reviews about restaurants,[13] excused eBay from claims by purchasers who bought forgeries from third party sellers,[14] and absolved Google from trademark liability for selling keywords as part of its advertising programme.[15] The basic spirit of two pieces of U.S. federal legislation enacted to establish a uniform

[8] See e.g., H.R. Rept. 105-551 Part I, at 11.
[9] *See* Christopher Cox, "The Origins and Original Intent of Section 230 of the Communications Decency Act", Richmond J of Law and Tech (Aug 27, 2020) <https://jolt.richmond.edu/2020/08/27/the-origins-and-original-intent-of-section-230-of-the-communications-decency-act/>. See also Robert Cannon, "The Legislative History of Senator Exon's Communications Decency Act: Regulating Barbarians on the Information Superhighway," 49(1) Federal Communications Law Journal, Article 3 (1996) <https://www.repository.law.indiana.edu/fclj/vol49/iss1/3>.
[10] See Lilian Edwards, *Role and Responsibility of Internet Intermediaries In The Field of Copyright and Related Rights* 2 (WIPO, 2010) <https://www.wipo.int/export/sites/www/copyright/en/doc/role_and_responsibility_of_the_internet_intermediaries_final.pdf>.
[11] Electronic Frontier Foundation, CDA 230: The Most Important Law Protecting Internet Speech, https://www.eff.org/issues/cda230/infographic.
[12] *Viacom Int'l, Inc. v. YouTube, Inc.* 676 F 3d 19 (2d Cir. 2012).
[13] *Hassell v. Bird*, 5 Cal.5th 522 (SC. Cal. 2018).
[14] *Gentry v. eBay, Inc.*, (2002) 99 Cal.App.4th 816, 121 Cal.Rptr.2d 703.
[15] *Google France SARL and Google v. Louis Vuitton Malletier SA.*, joined Cases C-236/08 to C-238/08.

federal policy for regulating the Internet has been propagated worldwide as national rules and regulations.[16] Simply put, the CDA and the DMCA have enabled the Internet that we know today.[17]

4. Recently, the CDA and the DMCA have been the subject of intense legislative scrutiny.[18] Questions have also been raised by the U.S. Supreme Court as to whether the CDA immunity aligns with its text and the business processes conducted by intermediaries.[19] It has been argued that these immunities are outmoded, as they were written in a prior Internet era.[20] In fact, new technologies in data aggregation and machine learning are empowering intermediaries to stretch these statutory immunities. To understand why, it is apposite to scrutinize the mechanics of the CDA and DMCA immunities before considering them against the backdrop of the increasing use of automation and AI by the intermediaries.

---

[16] The safe harbour provisions of s 512 of the DMCA have been enacted in various forms in Australia, Bahrain, Central America-Dominican Republic states, Chile, Columbia, the European Union, Morocco, Oman, Panama, the People's Republic of China, Peru, Singapore, South Korea and the United Kingdom. See Daniel Seng, 'The State of the Discordant Union, An Empirical Analysis of DMCA Takedown Notices' (2014) 18 Virginia Journal of Law and Technology 369, http://ssrn.com/abstract=2411915. Provisions similar to s 512 include Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), Arts. 12-14; ss 193A-193DE, Copyright Act (Cap. 63, Rev. Ed. 2006), People's Republic of China Regulation on Protection of the Right to Network Dissemination of Information of 2006, Arts. 20-23; Malaysia Multimedia Act. Provisions similar to s 230, CDA include s 26, Electronic Transactions Act (Cap. 88, Rev. Ed. 2011); India IT Act (2000), s 79; Australia Broadcasting Services Act 1992 (Cth), Schedule 5, Cl 91(1). For ease of discussion, reference will henceforth be made exclusively to the CDA and DMCA, although the equivalent national provisions in the respective jurisdictions should also be noted.

[17] See e.g., Ambika Kumar, The Test of Time: Section 230 of the Communications Decency Act Turns 20, Sep. 2016, https://www.dwt.com/blogs/media-law-monitor/2016/08/the-test-of-time-section-230-of-the-communications;; David Kravets, 10 Years Later, Misunderstood DMCA is the Law That Saved the Web 27, Wired, Oct. 27, 2008, https://www.wired.com/2008/10/ten-years-later/https://www.wired.com/2008/10/ten-years-later/.

[18] See e.g., Mark MacCarthy, Back to the future for Section 230 reform, Brookings Institute, https://www.brookings.edu/blog/techtank/2021/03/17/back-to-the-future-for-section-230-reform/ (noting that reform of the CDA is on the agenda for both the U.S. Congress and the Biden administration); Rebecca Tapscott, Senator Tillis Releases Draft Bill to Modernize the Digital Millennium Copyright Act, IP Watchdog, Dec. 22, 2020, https://www.ipwatchdog.com/2020/12/22/tillis-draft-modernize-dmca/id=128552/; Case C-401/19: Action brought on 24 May 2019 — Republic of Poland v European Parliament and Council of the European Union, 2019 O.J. (C 270) 21 (filing a legal challenge against the takedown-and-staydown notice rule in the Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130) 92 (May 17, 2019) (EU Copyright Directive) on the argument that it undermined the right of freedom of expression and was neither proportional nor necessary).

[19] See e.g., MalwareBytes, Inc. v. Enigma Software Group USA, LLC., 592 U.S. ___ , 141 S.Ct. 13 (2020) (criticizing the reading of extra immunity into s 230, CDA, per Thomas J.).

[20] See e.g., Matthew G Jeweler, "The Communications Decency Act of 1996: Why § 230 is Outdated and Publisher Liability for Defamation Should be Reinstated Against Internet Service Providers" (2008) 8 Pittsburg Journal of Technology Law & Policy 40.

The Mechanics and Limits of CDA Immunity

5. The immunity rule in section 230(c) of the CDA reads:

> (1) Treatment of publisher or speaker: No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.

6. The equivalent provision in the Singapore Electronic Transactions Act reads:[21]

> 26.—(1) Subject to subsection (2), a network service provider shall not be subject to any civil or criminal liability under any rule of law in respect of third-party material in the form of electronic records to which he merely provides access if such liability is founded on —
>
> (a) the making, publication, dissemination or distribution of such materials or any statement made in such material; or
>
> (b) the infringement of any rights subsisting in or in relation to such material.

7. Both the CDA and the ETA posit an immunity for an intermediary operating as a "service provider" for third party content.[22] The pithy formulation in the CDA grants immunity to an intermediary only if, as "an interactive computer service" provider,[23] it is not also an "information content provider", defined as "a person or entity that is responsible, in whole or in part, for the *creation or development* of information provided through the Internet or any other interactive computer service".[24] With this rule, resolution of the immunity turns on characterizing the intermediary as either a "content provider" of third party content, who has no immunity, or as a "service provider", who has immunity.[25]

8. But this simple distinction shades into penumbras of uncertainty with the advent of Web 2.0. Unlike their Web 1.0-era counterparts, wikis, blogs, social networks, podcasts, and interactive

---

[21] Cap. 88, Rev. Ed. 2011 (ETA).
[22] The Singapore formulation does not define what constitutes a "network service provider", as does the German formulation from which the Singapore provision is taken. See e.g., Ulrich Wuermeling, "The First National Multimedia Law - How Germany Regulates Online Services and the Internet" (1998) 14 Comp L & Sec Rep 41, 42.
[23] S 230(f)(2), CDA (defining an "interactive computer service" as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions").
[24] S 230(f)(3), CDA. Emphasis added.
[25] Zeran v. Am. Online, Inc., 129 F.3d 327, 333 (4th Cir. 1997).

websites emphasize user-generated content that features collaboration, contribution and participation from different users.[26] Web 2.0 upgrades the Web 1.0 experience by having an intermediary jointly create, co-opt or involve the user as the third party in the creation of some content. Would this heightened involvement of the intermediary displace its immunity? On this, the *en banc* majority of the 9th Circuit in Fair Housing Council of San Fernando Valley v. Roommates.com LLC said:

> A website operator can be both a service provider and a content provider: If it passively displays content that is created entirely by third parties, then it is only a service provider with respect to that content. But as to content that it creates itself, or is "responsible, in whole or in part" for creating or developing, the website is also a content provider. Thus, a website may be immune from liability for some of the content it displays to the public but be subject to liability for other content.[27]

9. So, an intermediary that designed its questionnaire, search and email systems to limit the listings available to subscribers based on sex, sexual orientation and presence of children became a content provider of such listings and did not have s 230 immunity, even though the answers were supplied by the advertisers.[28] This is because the intermediary's efforts contributed to the discrimination of tenants on the basis of their gender, family status, and sexual orientation in breach of the Fair Housing Act. Likewise, when an intermediary contracted with and paid researchers to obtain private telephone records and other confidential information which could only be obtained in breach of U.S. federal law, and then knowingly transformed the information into a publicly available commodity, it was responsible for the development of this specific content and would not be shielded by s 230.[29] On the other hand, an online dating site that offered neutral posting tools and did nothing to encourage the posting of defamatory content retained its s 230 immunity for defamation when a subscriber created a defamatory profile.[30] Likewise, when an intermediary developed a rating system by aggregating user-generated feedback or reviews, the intermediary was not treated as a content provider of the ratings and retained its s 230 immunity.[31]

---

[26] Wikipedia, Web 2.0, <https://en.wikipedia.org/wiki/Web_2.0>.
[27] 512 F.3d 1157, 1162-1163 (9th Cir. 2008).
[28] ibid 1169-70.
[29] FTC v Accusearch Inc 570 F.3d 1187 (10th Cir. 2009).
[30] Carafano v Metrosplash.com, Inc., 339 F.3d 1119 (9th Cir. 2003).
[31] Gentry v. eBay, Inc., 99 Cal.App.4th 816, 834, 121 Cal.Rptr.2d 703 (2002); Levitt v. Yelp! Inc., 2011 WL 5079526 (N.D. Cal. 2011); Kimzey v. Yelp! Inc., 836 F.3d 1263 (9th Cir. 2016).

10. What if an intermediary wishes to enforce content guidelines and review, or even remove, content that it considers objectionable? A less frequently applied part of s 230, known as the Good Samaritan provision, provides that the intermediary who undertakes these filtering duties or "self-regulation" shall not be held liable.[32] The Good Samaritan provision was enacted to reverse the controversial decision of Stratton Oakmont v. Prodigy Servs. Co.,[33] which held Prodigy liable in defamation for adopting content guidelines and filtering its subscribers' insulting and harassing postings. But while the aim behind this provision is laudable, most intermediaries steer clear of it in practice. This is because filtering its users' content nominally engages an intermediary with third party content, encourages a possible recharacterization of the intermediary as a content provider or developer of that content, and potentially weakens its possible reliance on s 230 immunity.[34]

The Rise of Automation and Machine Learning

11. Thus, whether the intermediary has immunity depends on whether it could be said to be responsible for the creation or development of content. This analysis is however based on cases involving intermediaries operating Web 2.0 websites. With technological advances like Web 3.0 and machine learning, the role of the intermediary has expanded beyond that of a service provider. In operating services such as aggregating, indexing, classifying, categorizing, formatting, enriching, and re-presenting user-originated content through targeted advertising or curated content to make for a more autonomous and intelligent Internet[35], intermediaries are moving away from their role as passive service providers and becoming "active" content delivery platforms. Is this permitted under the CDA immunity rules?

12. At first glance, this does not appear to be possible, since the intermediary has clearly taken on content creation or development responsibilities. But there is an escape route for intermediaries. In a tacit recognition of the increasing role that automation may play in serving online content, the Ninth Circuit opined that "[t]he mere fact that an interactive computer service 'classifies user characteristics ... does not transform [it] into a "developer" of the "underlying misinformation."'"[36] It ought to be noted that the Ninth Circuit was referring to its earlier decision in Carafano v

---

[32] S 230(c)(2), CDA.
[33] 1995 N.Y. Misc. LEXIS 229, 1995 WL 323710, 23 Media L. Rep. 1794 (N.Y. Sup. Ct. May 24, 1995).
[34] Barrett v. Rosenthal, 51 Cal. Rptr. 3d, 55, 70 (Cal. 2006). See also Jeweler, "The Communications Decency Act of 1996: Why § 230 is Outdated and Publisher Liability for Defamation Should be Reinstated for Internet Service Providers" (2008) 8 Pittsburg Journal of Technology Law & Policy 40, at 25-26; Sevanian, Andrew M, "Section 230 of the Communications Decency Act: A Good Samaritan Law without the Requirement of Acting as a Good Samaritan" (2014) 21 UCLA Entertainment Law Rev 121, 131-135.
[35] See e.g., Wikipedia, Semantic Web (Redirected from Web 3.0) <https://en.wikipedia.org/wiki/Semantic_Web#Web_3.0>.
[36] Roommates (n 27) 1173.

Metrosplash.com, Inc. where Metrosplash sought to summarize each user-submitted profile based on a Metrosplash questionnaire.[37] This summary was an attempt to identify similar profiles so that Metrosplash could provide matching services.[38]

13. But while one might excuse Metrosplash's content "input" as merely editorial and agree with the Ninth Circuit that this did not amount to the creation or development of new content, the characterization of intermediaries' involvement in other cases may lead to a number of questionable results. These involve situations where an intermediary uses automation to greatly expand its "editorial" role to arguably create or develop new and illicit content from user-supplied information, and yet avoid responsibility for "materially contributing" to the unlawful content. In doing so, intermediaries push the limits of the CDA immunity to its breaking point.

14. For instance, in *Jane Doe No. 1 v. Backpage.com*, three plaintiffs, all minors, sued Backpage.com, as each had been the subject of sex trafficking through advertisements placed on Backpage. They alleged that Backpage had facilitated this process by selectively removing postings made by victim support organizations and law enforcement sting advertisements. Backpage had tailored its posting requirements to make sex trafficking easier, including providing automated anonymization features such as message forwarding services and auto-replies (so that the advertisers could hide their e-mail addresses), and automatically removing metadata from uploaded photographs (so that they could not be scrutinized for their date, time and location). Backpage also allegedly crippled its automated filtering system, which would otherwise screen out advertisements with prohibited terms (so that advertisements with terms such as "brly legal" for "barely legal" and "high schl" for "high school" could still be posted), and accepted anonymous payments.[39] In what the court admitted was a "hard case", the First Circuit held that s 230 shielded Backpage from liability for participating in sex trafficking because these online features, "which reflect choices about what content can appear on the website and in what form, are editorial choices that fall within the purview of traditional publisher functions".[40]

15. Citing Metrosplash, the First Circuit ruled that Backpage was not an actual participant in a sex trafficking venture and was not complicit by merely using automated technical website designs and features.[41] But Backpage is clearly distinguishable: while both use questionnaires to collect information to create postings, Backpage took active steps to alter posting content, or coerce their

[37] Metrosplash (n 30) 1124.
[38] ibid.
[39] 817 F.3d 12, 16-17, 20 (1st. Cir. 2016).
[40] ibid 21-22.
[41] ibid 21.

modification, to shield its posters from easy identification. Backpage's obfuscation mechanisms were clearly associated with facilitating the illicit practice of sex trafficking, and could hardly be regarded as mere content-neutral editorial choices, while Metrosplash's categorization services resembled the table of contents or index pages of a publication. The analogy made is clearly unpersuasive.

16. The same reliance on the use of automation to preserve an intermediary's role as a mere "service provider" – and thus retain its s 230 CDA immunity – can be more clearly illustrated in Goddard v Google, Inc.[42] In this case, consumers brought a class action against Google for furthering a scheme whereby users were harmed when they clicked on web-based advertisements for fake mobile subscription services set up by third party advertisers through Google's AdWords advertising scheme. The U.S. District Court dismissed the class action. It ruled that Google's use of its AdWords "keyword tool" (which allowed advertisers to select keywords to correspond to their advertisements), and use of a mathematical algorithm as a "suggestion tool" (to suggest to advertisers the use of the word "free" in relation to "ringtone" to attract more mobile subscriptions), was a "neutral tool".[43] The court opined, without support, that Google "merely provides a framework that could be utilized for proper or improper purposes",[44] and the "selection of content was left exclusively to the [third party]."[45] In other words, automation – and even the use of AI-driven selection tools in Google's AdWords program that suggested content options to the third party – did not make the intermediary a "content provider". It was the third party who ultimately decided what content to use for its misleading and fraudulent advertisement.

17. The court supported this reasoning by contrasting AdWords with a Roommates scenario where it was suggested that a website that "remov[es] the word 'not' from a user's message reading '[Name] did not steal the artwork' in order to transform an innocent message into a libelous one" would void its CDA immunity.[46] That may be true where an intermediary converts a message into one with an entirely opposite meaning, but the analogy is incomplete. The court never considered the subtly persuasive – and ultimately coercive – power of machine-driven recommendation systems.[47] It is well known that Google AdWords operates, as a "self-service" product, one of the

[42] Goddard v. Google, Inc., 640 F.Supp.2d 1193 (N.D.Cal. 2009).
[43] ibid 1199. See also Hill v. Stubhub, Inc., 219 N.C.App. 227 (N.C. C.A. 2021).
[44] ibid.
[45] ibid 1197.
[46] ibid 1199.
[47] See e.g., Nick Sever, "Captivating algorithms: Recommender systems as traps" (2018) Journal of Material Culture 1 <https://www.anthropology.uci.edu/publications/Nick%20Seaver%20Journal%20of%20Material%20Culture.pdf>.

most sophisticated machine learning-powered bidding services worldwide.[48] It works by enabling the advertiser to make automated bids for keywords, based on the history of the advertiser, the history of the user, the relevance of the ad, the time and day when the auction is happening, and many other factors, to "deliver the most relevant ad to the user at the right moment for them."[49] Thus, when AdWords suggests the word "free" for an advertisement, it does so on the basis that the word would be the most relevant term the user is looking for. It ought to be noted that the word "free" in the context of Internet parlance has too often been associated with illicit activities,[50] just as it has also come to signify largely unjustified expectations of Internet consumers seeking advantageous deals online. It is clearly a "bait" word, which advertising systems have come to associate with greater online advertisement traction – and is associated with arguably ulterior intent when linked with mobile subscription services such as "ringtones" – because subscription services are inherently not "free".[51] In other words, if AdWords suggested the use of the word "free" with "ringtones", this disclosed possible complicity on the part of the intermediary. This issue needs to be investigated further and ought not be cursorily dismissed. Otherwise, the advent of AI and automation will enable an intermediary, as a "service provider", to erect "decisional firewalls" between itself and the offerings of its programmed systems. If unchallenged, Goddard v Google would suggest that an intermediary can retain its CDA immunity by ostensibly leaving the ultimate decision in the hands of the third-party user, who may be guided by a machine learning system programmed by the intermediary itself.

18. A similar case can be found in Force v Facebook, Inc., where victims of Palestinian attacks in Israel brought actions against Facebook for knowingly hosting accounts belonging to Hamas, classified by the US as a terrorist organization, contending that Facebook's social matching algorithms promoted terrorist content to people who liked similar pages or posts.[52] A majority of the

---

[48] Jerry Dischler, Putting machine learning into the hands of every advertiser, Google, Jul. 10, 2018, <https://support.google.com/google-ads/answer/9065075?hl=en>.
[49] Odolena Kostova, "Machine Learning, Smart Bidding and Google Ads", Medium, Feb. 5, 2019, <https://medium.com/@odolenakostova/machine-learning-smart-bidding-and-google-ads-1724aa8c9232>.
[50] *See* Anti-piracy code of practice for search engines proposed by rights holder representatives, Pinsent Masons, Jan. 27, 2012, <https://www.pinsentmasons.com/out-law/news/anti-piracy-code-of-practice-for-search-engines-proposed-by-rights-holder-representatives>. Google's Advertising Policies Help specifically desists from the use of words such as "free", which it claims are gimmicky and do not meet editorial and professional requirements. However, in Goddard, evidence was presented that the AdWords system offered the advertisers the use of the word "free" for mobile subscription services. See Google Ads policies, <https://support.google.com/adspolicy/answer/6008942?hl=en>.
[51] These are known as "negative keywords" because respectable advertisers run their ad campaigns to *avoid* users who conduct searches using these keywords. See e.g., Stephanie Mialki, How to Find, Add & Use Negative Keywords to Your Best Advantage, Jan. 7, 2020, <https://instapage.com/blog/negative-keywords>.
[52] Force v Facebook, Inc., 934 F.3d 53 (2nd Cir., 2019); Supreme Court cert. not granted.

Second Circuit dismissed the claims, holding that Facebook did not "develop" the terrorism-related content on its social networking site, by merely developing algorithms that use its users' information to match the "materially unaltered" content with other users.[53] In a clear recognition of the weakness of the majority's argument, Chief Judge Katzmann penned a strong dissent, noting how Facebook's algorithms had played a crucial role in fostering a unique global community by linking and engaging individual users through suggesting connections to other users with shared interests, in this case, in terrorism.[54] Chief Judge Katzmann found that Facebook had, through its social networking service, become a publisher, not of the users' content, but of the users' information, taking it out of the CDA immunity.[55] After all, "the creation of social networks [through matching algorithms] goes far beyond the traditional editorial functions that the CDA immunizes."[56] It is worth noting that Facebook does use AI and machine learning and manual reviewers to filter out offensive postings and content on Facebook and Instagram, in compliance with its Community Standards,[57] and this has involved an expenditure of considerable costs and resources. But Facebook did not rely on this to mount a Good Samaritan defence to the claims.

A final – and perhaps harder – example can be drawn from Yelp, Inc. Yelp's business model involves the collation and subsequent curation of recommendations and reviews about businesses. Yelp also runs a paid advertisement programme on the side to allow subscribers to promote their businesses. Because the raison d'être for Yelp is the hosting of third-party reviews, Yelp has to take steps to verify these reviews to protect its business model as a trustworthy source of reviews,[58] and it is therefore well-known that Yelp actively curates and controls the presentation of reviews.[59] However, Yelp has also been dogged by allegations of Yelp-manufactured negative reviews or wrongful manipulation of third-party reviews to the detriment of businesses who refuse to purchase

---

[53] ibid 70.
[54] ibid 82-83.
[55] ibid 82.
[56] ibid 83.
[57] Facebook, How enforcement technology works, Jun. 23, 2021, <https://transparency.fb.com/enforcement/detecting-violations/how-enforcement-technology-works/>. *See* also Facebook, ibid 60.
[58] This is to address the problem of "astroturfing", which is the practice of masking the sponsors of a message or organization to make it appear as though it originates from and is supported by grassroots participants. See Wikipedia, Astroturfing, https://en.wikipedia.org/wiki/Astroturfing. See also Neal Ungerleider, FTC Subpoena Revelations, Thousands Of Complaints Send Yelp's Stock Price Tumbling, Fast Company, Apr. 4, 2014, <https://www.fastcompany.com/3028725/ftc-subpoena-revelations-send-yelps-stock-price-tumbling>.
[59] The issue was first brought to the mainstream media by Wall Street Journal. See Angus Loten, "Yelp Regularly Gets Subpoenas About Users", The Wall Street Journal, Apr. 2, 2014, <https://www.wsj.com/articles/SB10001424052702303847804579477644289822928>. For instance, many businesses pay third party reviewers to flood their Yelp online listings with good reviews. See e.g., Curry v. Yelp Inc., 2015 WL 1849037, at *1 (N.D. Cal. 2015).

advertising from Yelp.[60] While these allegations have not been proved, it is known that Yelp enlists sophisticated recommendation software that could filter and curate reviews for their authenticity, quality and integrity,[61] and even automatically republish the curated reviews on search engines.[62] Such is the utility of automated curation that in the absence of actual proof of human intervention in the curation process, the automated nature of its editorial operations has allowed Yelp to successfully rely on the s 230 immunity to defend itself in various claims.[63] In this regard, while U.S. courts have noted that Yelp's machine-powered curation of reviews for subsequent publication could represent an immunized activity for filtering objectionable reviews and potentially qualify for Good Samaritan immunity,[64] they do not give these arguments much credence because to qualify, the intermediary has to demonstrate that the filtering was done in "good faith".[65] In contrast, there is no such limitation to acquire s 230 immunity.[66] For this reason, intermediaries like Yelp (and even Facebook[67]) seem to rely on expanding their services and their role as "developers" of user-supplied content and pushing the envelope of s 230 immunity, rendering any reliance on the Good Samaritan defence otiose.

The DMCA and Copyright Liability

19. In the copyright claims space, the same immunity that would apply to Internet intermediaries finds expression in a slightly different form in the DMCA, which seeks to codify the basic rule set out in *Religious Technology Centre v. Netcom On-Line Communication Services Inc.*[68] The main difference is that unlike s 230 of the CDA, the DMCA immunity for the four designated classes of Internet intermediaries against both direct and indirect copyright infringement is conditional, that is, granted subject to compliance with certain conditions (hereinafter referred to the "safe harbours"). With respect to intermediaries such as hosting and information location tool service providers,[69] the immunity is granted only if, among other conditions, the intermediary has no

[60] See e.g., Levitt v. Yelp! Inc., 2011 WL 507952, at *1 (N.D. Cal. 2011); Kimzey v. Yelp! Inc., 836 F.3d 1263 (9th Cir. 2016).
[61] See e.g., Curry (n 59) *1.
[62] See e.g., Kimzey (n 60) 1270.
[63] *See* above.
[64] The same argument could have influenced the decision of the majority in Force v Facebook, which did refer to the Good Samaritan protections in s 230(c)(2). *See* Force (n 52) 80.
[65] Levit (n 60) at *10.
[66] ibid.
[67] Force (n 52) 80.
[68] Netcom (n 5). The s 230 CDA immunity does not apply to matters pertaining to intellectual property claims. 47 U.S.C. § 230(e)(2). The same rule applies in Singapore. *See* s 26(2)(d), ETA.
[69] The two safe harbour defences that are of general application are s 512(c) (hosting service providers) and s 512(d) (information location service providers). The other two safe harbour defences relate to Internet

actual knowledge of infringement or responds expeditiously to a DMCA-prescribed takedown notice submitted by an aggrieved copyright owner, content provider or its agent (referred to as the reporter) to remove or disable access to the infringing material.[70] In other words, unlike s 230 of the CDA, DMCA immunity requires intermediaries to cooperate with content providers,[71] although the onus remains on the content provider to detect and report infringing materials online to the intermediary.

Volition and the Advent of Automation

20. DMCA immunity is, however, explicitly stated to operate without prejudice to existing defences in the law of copyright.[72] Furthermore, the conditional nature of DMCA immunity incentivizes an intermediary to shape its business such that it attracts neither direct liability – that is, liability for infringing conduct that the intermediary itself undertakes – nor indirect or secondary copyright liability – that is, accessory liability for illicit conduct undertaken by third parties. An intermediary is able to do so by adopting a business model that relies on automation to shift responsibility to the third party-user for any activity undertaken with copyright material.[73]

21. This is best illustrated with a series of cases that litigated the legality of network digital video recording (NDVR) services. Also known as remote storage digital video recorder (RS-DVR) services, this is a network-based digital video recorder (DVR) service where instead of storing media content, typically free-to-air public broadcast television content, on a DVR or set-top box at the consumer's private home, the content is stored in the cloud or on servers controlled by the intermediary service provider.[74] The recorded content is typically only available to the user who recorded it.

22. At first sight, the NDVR services appear to be an unobjectionable extension of time-shifting of broadcast programs, which, pursuant to the seminal decision of the U.S. Supreme Court in *Sony Corp. of America v. Universal City Studios, Inc.* has been held to be fair use of the content, since the recording is by the user for his private and domestic use.[75] (Time-shifting has been statutorily

---

intermediaries as transitory digital network communications service providers (s 512(a)) and service providers providing system caching (s 512(b)).

[70] S 512(c)(1)(A); 512(d)(1)(A). Other conditions include appointing a designated agent to receive notifications (s 512(c)(2)), implementing a repeat infringer policy (s 512(i)(1)(A)) and accommodating and not interfering with standard technical measures (s 512(i)(1)(B)).

[71] Edwards (n 10) 6.

[72] S 512(l).

[73] CoStar Group, Inc. v. LoopNet, Inc., 373 F.3d 544, 554 (2004), quoting from ALS Scan, Inc. v. RemarQ Communities, Inc., 239 F.3d 619, 622 (4th Cir.2001). S*ee also* Daniel Seng, "Detecting and Prosecuting IP Infringement with AI: Can the AI Genie Repulse the Forty Counterfeit Thieves of Alibaba?" *Artificial Intelligence and Intellectual Property* (Lee, Hilty and Liu eds, OUP 2021) 292.

[74] Wikipedia, Network DVR, <https://en.wikipedia.org/wiki/Network_DVR>.

[75] 464 U. S. 417 (1984).

sanctioned in the copyright laws of many countries.[76]) However, broadcasters have objected to NDVR services on the basis that use of the intermediaries' services encroaches on their exclusive right to transmit (and retransmit) content. Thus, one of the preliminary issues that must be resolved is whether the making of the NDVR copies and the subsequent transmission of these recorded copies using the intermediary's platform is done by the *user* or by the *intermediary*.[77]

23. As the late Scalia J explained in his powerful dissent in American Broadcasting Cos., Inc. v. Aereo, Inc., the difference turns on whether the making of the copies and their subsequent transmission is considered the product of the user's or the product of the intermediary's "volitional conduct".[78] Even though copyright infringement is based on strict liability, "there should still be some element of volition or causation which is lacking where a defendant's system is merely used to create a copy by a third party."[79] Drawing upon the analogy that the owner of a copy machine is not considered to be a direct infringer if a customer uses the machine to duplicate an infringing work, the U.S. Courts of Appeals have uniformly[80] concluded that this requires courts to identify the "actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner."[81] This reasoning is not exclusive to U.S. case law. The Australian High Court in Roadshow Films Pty Ltd v. iiNet Ltd acted on a similar basis when it concluded that there was no "reasonable basis"[82] for the intermediary to take action to terminate the accounts of its subscribers alleged to have used BitTorrent file sharing software, or that it was "not unreasonable"[83] for the intermediary to not do so. The High Court noted that an intermediary was not held liable "merely because" it has provided facilities for enabling the infringement by the user who is the primary infringer.[84] This parallels the observation of the Ninth Circuit that establishing volition is, in the language of proximate causation, simply showing that the conduct in question is the "*direct* cause of the infringement."[85]

---

[76] E.g., s 114, Singapore Copyright Act; s 111, Australian Copyright Act. For a comparative analysis on time-shifting laws, see Van Goethem, Arvind, "A Comparative Analysis on the Legality of Cloud Personal Video Recorders" (November 17, 2015), <https://ssrn.com/abstract=2729801>.
[77] See e.g., American Broadcasting Cos., Inc. v. Aereo, Inc., 573 US 431, 453 per Scalia J (2014).
[78] ibid 456.
[79] Netcom (n 5) 1370.
[80] See Aereo (n 77) 453 per Scalia J.
[81] CoStar (n 73) 550.
[82] [2012] HCA 16, [78] per French CJ, Crennan and Kiefel JJ.
[83] ibid [146] per Gummow and Hayne JJ.
[84] ibid [136] per Gummow and Hayne JJ.
[85] Perfect 10, Inc. v. Giganews, Inc., 847 F.3d 657, 666 (9th Cir. 2017), quoting from Perfect 10 Inc., v. Giganews, Inc., 2014 WL 8628034, at *7 (C.D. Cal. Nov. 14, 2014) (emphasis in the original).

24. Thus, while the intermediary did indeed build the automated system for making NDVR recordings, "the key point is that subscribers call all the shots", since the automated system could not make any recording or relay any recording until the subscriber selected the programme he wanted and requested that it be relayed.[86] The Second Circuit in The Cartoon Network LP, LLLP v CSC Holdings, Inc.[87] and the Singapore Court of Appeal in RecordTV Pte Ltd v MediaCorp TV Singapore Pte Ltd[88] also arrived at the same conclusion. As the Court of Appeal in RecordTV opined:

> [S]ince the only [the content provider's] shows that were "communicated" were those shows that appeared on each Registered User's playlist, and since the exact make-up of each playlist depended on the specific shows which the Registered User in question had requested to be recorded, "the person responsible for determining the content of the communication at the time the communication [was] made" would be that Registered User himself. [The intermediary] would not have been the communicator of the [content provider's] shows …[89]

25. It ought to be noted that the issue of "volitional conduct" is not always resolved in favour of the intermediary. In National Rugby League Investments Pty Limited v Singtel Optus Pty Ltd, the Full Court of the Federal Court of Australia held that the time-shifted copy was made by the NDVR intermediary service provider, or alternatively, by both the service provider *and* the subscriber.[90] Likewise, the Japanese Supreme Court held in a pair of decisions – Maneki TV[91] and Rokuraku II[92] – that it was the intermediary service provider who was responsible as it had developed the environment for making it uncomplicated and almost effortless to make the reproductions and the retransmissions. "But for such actions carried out by the service provider", the Japanese Supreme Court noted, it would not be possible for users to record and reproduce the broadcast programmes.[93] In American Broadcasting Cos., Inc. v. Aereo, Inc. itself, the U.S. Supreme Court majority affirmed the illegality of Aero's NDVR services, though it sidestepped the issue of volition by simply concluding that because the services provided by the Intermediary resembled cable-TV, it

[86] Aereo (n 77) 456 per Scalia J.
[87] 536 F 3d 121, 131-132 (2nd Cir, 2008).
[88] [2011] 1 SLR 830, [15], [34].
[89] ibid [36].
[90] [2012] FCAFC 59, [5].
[91] 2009 (Ju) No. 653; Minshu Vol. 65, No. 1 (Japanese Supreme Court, Jan. 18, 2011).
[92] 2009 (Ju) No. 788; Minshu Vol. 65, No. 1 (Japanese Supreme Court, Jan. 20, 2011).
[93] ibid.

ought to be regulated as such,[94] a result which Scalia J chastised as a "result-driven"[95] rule that "provides no criteria for determining when its cable-TV-lookalike rule applies."[96]

26. In summary, the line of cases relating to the use of automation to provide online services to users, such that the infringing activities committed by the users could not be ascribed to the service provider's "volitional conduct", has been met with a mixed degree of success. Certainly, service providers in these cases have met with less success in shifting responsibility to users, and preserving their immunities under the DMCA, than their CDA counterparts. This phenomenon can be explained by two factors: first, the copyright jurisprudence on "volitional conduct" relies less on the form taken by the service provider's automation of its services and more on the substance of these services. Second, DMCA safe harbours operate as conditional immunities without prejudice to existing and more flexible rules of copyright and tortious causation. There are, however, additional issues triggered by the use of automation with respect to the operation of the DMCA safe harbours.

Automated Processing, Errors in Takedown Notices and the Imputation of Bad Faith

27. As previously noted, the DMCA safe harbours operate as conditional immunities, which require an intermediary to act expeditiously on an effective takedown notice. An effective takedown notice is one that complies with the six statutory requirements prescribed for a notice – (i) an authorized signature, (ii) description of the copyrighted work, (iii) identification of the material claimed to be infringing – also known as the takedown request, (iv) the takedown reporter's contact information, (v) a statement of good faith belief that use of the material complained of is not authorized, and finally, (vi) a statement of accuracy as to the information in the notice and confirmation that the reporter is authorized by the copyright owner or exclusive licensee.[97] The DMCA goes on to provide that exact compliance with these formalities is not required – only substantial compliance is required.[98] Nonetheless, if there is no substantial compliance with formalities (ii), (iii) and (iv), the notice will fail *in limine* and the intermediary is entitled to disregard it as being erroneous.[99]

---

[94] Aereo (n 77) 442-444 per Breyer J.
[95] ibid 461 per Scalia J.
[96] ibid 460 per Scalia J. The majority did not explicitly repudiate Scalia J's formulation of the volitional-conduct requirement. See BWP Media USA, Incorporated v. T & S Software Associates, 852 F.3d 436, 441 (5th Cir. 2017).
[97] S 512(c)(3)(A)(i) to (vi). See also Singapore Copyright (Network Service Provider) Regulations 2005, Rg 7, regulation 3.
[98] S 512(c)(3)(A). See also Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1112 (9th Cir. 2007).
[99] The DMCA obliges an intermediary to provide the reporter with a second chance to remedy defects in formalities (i), (v) and (vi). Daniel Seng, Copyrighting Copywrongs: An Empirical Analysis of Errors with Automated DMCA Takedown Notices, 37 Santa Clara High Tech. L. J. 119, 138 (2021).

28. This is because formalities (ii) and (iii) enable the intermediary to identify the infringed work and the infringing material,[100] and formality (iv) enables the reporter to be contacted in the event a counter-takedown notice is served on the intermediary.[101] Yet, surprisingly, data from empirical studies conducted on takedown notices show that a significant number of takedown notices do not have formalities (ii) and (iii). For instance, notices with no copyright work descriptions account for up to 9.6% of all notices issued[102] in 2013 before dropping to a negligible 0.05% of all notices in 2015.[103] However, notices with no takedown requests continue to make up a substantial number of all notices, rising to 12.4% of all notices in 2013 and 11.4% of all notices in 2014 before falling slightly to 7.3% of the notices in 2015.[104]

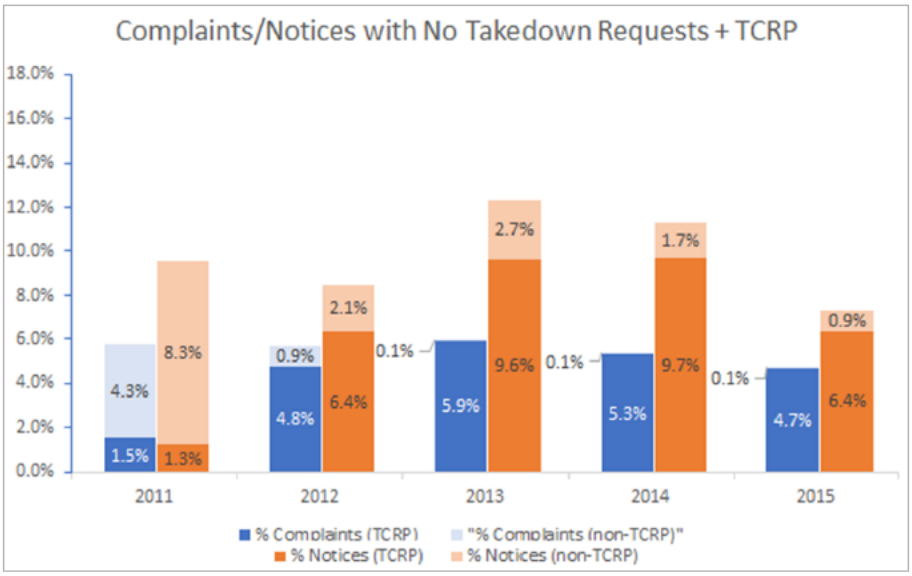

Figure 5: Plot of Percentage of Notices and Complaints with No Takedown Requests (by TCRP Reporters) between 2011 and 2015[186]

Figure 1: Chart comparing Error Rates of Automatically- vs Manually-Processed Notices and Complaints

29. The empirical research shows that most of these erroneous notices with no takedown requests are issued by Google's Trusted Copyright Removal Program ('TCRP') reporters - reporters who are considered more trustworthy and are empowered, almost exclusively, to use automated means to submit takedown notices to Google Inc.[105] These are known in the industry as "robo

[100] ibid 139-140.
[101] S 512(g)(2)(B).
[102] Notices issued and recorded in the Lumen database.
[103] *Copyrighting Copywrongs* (n 99) 154 (Table 4).
[104] ibid.
[105] ibid, at 159.

takedowns".[106] In contrast, the number of erroneous notices by non-TCRP reporters is several orders of magnitude smaller, as the figure above shows. One possible hypothesis is that these errors are an inevitable byproduct of automated enforcement: when content providers and their reporters use automated means to detect instances of online infringement and report them to Internet intermediaries like Google and Twitter, automation involves a tradeoff between accuracy and efficiency.[107] However, as the empirical research also shows, TCRP reporters have widely varying rates of such errors.[108] In fact, some of the top takedown notice reporters (by volume of takedown notices), such as Stichting BREIN, AudioLock.NET, Degban, and RIAA, have the smallest ratio of empty notices to total notices.[109] This is clearly indicative of process errors on the part of the poorly-performing reporters – errors in the design and configuration of their automated reporting systems.[110]

Takedown notices are also not immune to substantive errors – errors that raise substantive legal questions that undermine the underlying claim for alleged copyright infringement. One example of a substantive error is a "spent" takedown request: a takedown notice targeting a website which is no longer functional – even though the claim asserts that it is valid and the information in the notice is accurate.[111] While the number of these "spent" requests is small – an empirical study suggests that *one type* of "spent" requests[112] accounts for only 0.23% of all takedown requests – their absolute number is not small. All in, 2.74 *million* clearly invalid requests have been issued between 2011 and 2015[113] – requests which need not be attended to by the intermediary or which would affect its DMCA immunity, but which unnecessarily consume the intermediary's resources in acting on and responding to them.

30. The DMCA *does* provide for penalties against a reporter who "knowingly materially misrepresents" that material or activity is infringing.[114] Courts are beginning to recognize the dangers of having these bad notices and requests overwhelm Internet intermediaries, and there have been rulings that the issuance of defective takedown notices may be grounds for the

[106] Daniel Seng, "The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices", 18 Va. J.L. & Tech. 369, 398-400 (2013); Zoe Carpou, Robots, Pirates, and the Rise of the Automated Takedown Regime: Using the DMCA to Right Piracy and Protect End-Users, 39 Colum. J. L. & Arts 551 (2016).
[107] *Copyrighting Copywrongs* (n 99) 165.
[108] ibid 161.
[109] ibid 164.
[110] ibid 162-163.
[111] S 512(c)(3)(A)(v), (vi).
[112] Based on the Megaupload test. See *Copyrighting Copywrongs* (n 99) 171-182.
[113] ibid 181.
[114] S 512(f).

Intermediary to mount an action for knowing misrepresentation.[115] However, there are difficulties in making such claims against the notice reporters because the damage suffered by the intermediary must be proved.[116]

31. More critically, the misrepresentation can only be constituted as a "knowing misrepresentation" if it is proved that the reporter "should have known [about and not issued the notice or request] if [they have] acted with reasonable care or diligence or would have had no substantial doubt had it been acting in good faith."[117] Thus puts a very high bar on the aggrieved intermediary, especially if the reporter pleads that the errors are the result of a misconfiguration of its automated technical processes. It could even mount a plausible argument that the errors were driven by out-of-control machine-learning algorithms, and that these were outliers in the programmed space of the system's operations. The issue of tortious liability for out-of-control software agents has been explored elsewhere,[118] and this author takes the view that this is a smokescreen argument that should not detract from the conclusion that the ultimate causality of these errors is still the reporter, with a misconfigured algorithm that was under its control.[119] It suffices to say that the DMCA threshold to make a successful claim for material misrepresentation is set too high to make misrepresentation claims a real incentive for reporters to verify their takedown notices and report their claims correctly and accurately.[120]

32. If this problem is not remedied, the increasing number of notice mistakes made by reporters, coupled with the DMCA conditions and the difficulty of prosecuting reporters for these mistakes, will force intermediaries into a state of disregard. This is exactly what is happening now, with some intermediaries reporting extremely high rates of successful takedowns notwithstanding the formal and substantive mistakes made by reporters.[121] Part of the reason could be that these high rates of "successful" takedowns arose because the intermediaries themselves had been deploying automation and machine learning to deal with the ever-increasing volume of takedown notices, complaints, and requests received.[122] While automation and AI have enabled intermediaries

---

[115] See e.g., Rosen v. Hosting Services, Inc., 771 F. Supp. 2d 1219 (C.D. Cal. 2010).
[116] See e.g., Lenz v. Universal Music Corp., No. 5:07-cv-03783-JF, 2013 WL 271673, at *9 (N.D. Cal. Jan. 24, 2013; Automattic Inc., v. Steiner, 82 F. Supp. 3d 1011, 1030 (N.D. Cal. Mar. 2, 2015).
[117] Online Policy Group v. Diebold, Inc., 337 F. Supp. 2d 1195, 1204 (N.D. Cal. 2004).
[118] See e.g., Daniel Seng and Tan Cheng Han, AI and Agents.
[119] The empirical data suggests that this argument is implausible, because there were many reporters who submitted takedown requests that do pass the Megaupload test. *Copyrighting Copywrongs* (n 99) 181.
[120] ibid 186.
[121] Google, for instance, reports a successful takedown rate of 97.5% from 2011 to 2012, and above 98% in 2015, and Microsoft a successful takedown rate of 99.7%. ibid 126.
[122] See Daniel Seng, "'Who Watches the Watchmen': An Empirical Analysis of Google's Rejected Copyright Takedown Notices" (2015) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3687861>.

to scale up their processing of takedown notices, their unchecked use has created an environment that lacks transparency and accountability, resulting in opportunities for misuse and abuse.[123]

Reform

33. As intermediaries continue to accrete and add new services, their influence over the content that users see or receive also increases. The immunity laws in the CDA and DMCA may represent in the Web 1.0 era the correct balance between protecting intermediaries from third-party content and requiring them to serve as gatekeepers to shield Internet users from illicit and illegal content and information. But automation and AI technologies today threaten to upend this delicate balance. The use of machine learning to format and present information empowers the intermediary to control and shape such information, while immunizing it because the user is the content developer. This in turn emasculates the Good Samaritan provision and discourages any intermediary from discharging its gatekeeping responsibilities. Likewise, under the DMCA, intermediaries use automation to shift responsibility for content to end users and preserve their copyright immunity, and fail to take a rigorous approach towards filtering out erroneous takedown notices.

34. The Internet of the future will be ever more all-encompassing and more customized,[124] and our reliance on intermediaries will be even greater. To extract the most from this increasingly vital and all-encompassing platform, we want to preserve intermediary neutrality and minimize the chilling effect of censorship and content regulations. Yet at the same time, we need intermediaries to keep our Internet safe, reliable and trustworthy.

35. The recently proposed EU Digital Services Act[125] is the latest attempt to rebalance these rules. Like the CDA, it confirms the horizontal intermediary immunity for third-party content.[126] But like the DMCA, it also requires intermediaries (including OPs)[127] to set up a mechanism to receive, from any reporter, including "trusted flaggers",[128] notices seeking to takedown illegal content.[129]

---

[123] Copyrighting Copywrongs (n 99) 185.
[124] See e.g., Matt Blitz, What Will the Internet Be Like in the Next 50 Years?, Popular Mechanics (Nov. 1, 2019), <https://www.popularmechanics.com/technology/infrastructure/a29666802/future-of-the-internet/>.
[125] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, Dec. 15, 2020 (DSA).
[126] ibid, Preamble (17)-(18); arts. 3-5, 7.
[127] The DSA identifies three types of intermediary services: conduits, caching and hosting services. ibid, art. 2(f). Online platforms (OPs) are a type of hosting service that disseminate the hosted content to the public. ibid, art. 2(h), and "very large online platforms" (VLOPs) have additional obligations, including risk assessments and risk mitigation measures. ibid, arts. 26, 27. Intermediaries referred to here are hosting services.
[128] ibid, art. 19 (entities designed by the Digital Services Coordinator that have expertise, represent collective interests and are diligent).
[129] ibid, art. 14. See also arts. 15 (reasons for takedown), 17 (putback).

Intermediaries may optionally[130] implement their own "content moderation" by way of human review or algorithmic decision-making[131] to identify and remove illegal content.[132] To promote transparency, intermediaries have to provide and publish in a database the reasons for disabling content.[133] (Under the DMCA, some intermediaries also publish takedown notices to the Lumen database.[134]) They also have to publish a comprehensible and detailed report of the content moderation undertaken[135] (with OPs also required to publish the specification, accuracy and safeguards of the automated means used).[136]

36. To promote accountability in the use of the takedown notices, where a "trusted flagger" has submitted a significant number of insufficiently precise or inadequately substantiated notices, it may be suspended[137] or even lose its trusted status.[138] This last rule mirrors a proposal first made by the author in 2015 for the publication of "accountability metrics" so that reporters who repeatedly make mistakes (through their robo-takedown systems) will have the priority of their notices downgraded.[139] Regrettably, the DSA did not mandate another proposal made in 2015 that reporters have to first validate the URLs that they are seeking to disable. As shown above, validating the URLs does not require considerable resources, but doing so would greatly enhance the accuracy and trustworthiness of takedown notices.[140]

37. In a first, the DSA regulates the use by VLOPs[141] of machine learning "recommender systems" that suggest specific information to recipients of the service,[142] and "advertisement" systems that promote messages or information.[143] VLOPs have to put in place risk assessment[144] and risk mitigation measures[145] and be assessed for compliance by an independent auditor,[146] whose

[130] ibid, art. 13(1)(c).
[131] ibid, art. 12(1).
[132] ibid, art. 2(p).
[133] ibid, art. 15.
[134] Lumen, About Us, <https://www.lumendatabase.org/pages/about>.
[135] DSA, art. 13. Intermediaries like Google, Facebook and Microsoft already publish semiannual "transparency reports".
[136] ibid, arts. 23(1) (OPs), 33 (VLOPs).
[137] ibid, art. 20(2).
[138] ibid, art. 19(5)-(6).
[139] Copyrighting Copywrongs (n 99) 186-188.
[140] ibid.
[141] DSA, art. 25 (defined as intermediaries that service at least 45 million active recipients).
[142] ibid, art. 2(o).
[143] ibid, art. 2(n).
[144] ibid, art. 26.
[145] ibid, art. 27.
[146] ibid, art. 28.

report shall be publicly available.[147] In addition, VLOPs have to disclose the "main parameters" of their "recommender systems" and enable recipients to modify or influence these parameters.[148]

38. The DSA is certainly to be lauded for rules that promote the transparent use of AI by intermediaries. It however immunizes rather than holds accountable the intermediaries' use of automated systems to aggregate and disseminate illicit content that target specific groups or trigger specific harms. While the DSA envisioned the intermediaries' use of content moderation to remove such content, it does not mandate the intermediaries' own content moderation or condition the immunities on their use.[149] In fact, because there is no obligation to monitor,[150] there is no incentive for any intermediary to conduct content moderation. As this study suggests, the increasing use of automation and AI means that intermediaries are more, not less, likely to rely on the immunities to justify the use (and abuse) of their services.

39. Any impactful reform must explicitly recognize the developing role of automation and machine learning systems in the services offered by Internet intermediaries. And the immunities must be concomitant with adequate accountability, such that intermediaries are obliged to minimize harmful or illicit content. For starts, the intermediary's immunity is not absolute: even under s 230, intermediaries may be liable for crimes relating to the sexual exploitation of children and federal criminal statutes, breaches of intellectual property, communications privacy and sex trafficking laws.[151] There is a baseline of third-party activities and content which intermediaries have to guard against. This translates into a minimum obligation to monitor and guard their platforms, which intermediaries can, should and have deployed content moderation systems to filter and remove such content.[152] Thus, the assertion that "there is no general monitoring obligation or active fact-finding obligation"[153] must be heavily circumscribed. To this end, it is further proposed to condition the immunity on the intermediary's good faith discharge of basic content moderation. While this deviates from provisions in the DMCA and DSA,[154] it is the only solution to bring the Good Samaritan rules to bear by incentivizing the intermediaries to bring some order to the unruly online

[147] ibid, art. 33(2). Confidential information or information that may cause significant vulnerabilities or undermine public security or harm recipients may be removed. ibid, art. 33(3).
[148] ibid, art. 29(1).
[149] ibid, arts. 6, 13.1(c).
[150] ibid, arts. 6, 7.
[151] S 230(e), CDA.
[152] DSA, preamble (58). See also art. 26(1).
[153] DSA, art. 7; cf. DMCA, s 512(m)(1).
[154] ibid.

environment that they have engendered. Indeed, this change would make the immunities even more relevant and pertinent to all parties in an increasingly complex digital world.

Conclusion

40. As Kranzberg observed in the early part of the 20th century, "[t]echnology is neither good nor bad, nor is it neutral."[155] The same adage can quite aptly be applied to AI and its machine learning implementations. It is up to us to infuse technology with the best of our human values. If one such value is that we should "do no harm",[156] it should be observed regardless of the environment, physical or virtual, or our technological tools. When Tim Berners-Lee first proposed the web, he envisioned it as a pool of information that would grow and evolve, as we grow and evolve.[157] It is therefore hoped that any recalibration to the law regarding intermediary liability will therefore be imbued with the same wisdom and foresight that made possible the Internet in the first place, so that it can continue to flourish as the greatest heritage of our human civilization.

(9503 words, footnotes included, excluding abstract)

[155] Kranzberg's First Law. *See* Eric Schatzberg, Kranzberg's First and Second Laws, Dec. 20, 2018, <https://www.technologystories.org/first-and-second-laws/>.
[156] John Stuart Mill, On Liberty (1859).
[157] Tim Berneres-Lee, Information Management: A Proposal, Mar. 1989, <https://www.w3.org/History/1989/proposal.html>.