

CDA 230 FOR A SMART INTERNET

Madeline Byrd* & Katherine J. Strandburg**

INTRODUCTION

Section 230 of the Communications Decency Act of 1996¹ (“CDA 230”), which provides broad-based immunity to online service providers for third-party content, has been hailed as “the law that gave us the modern internet.”² Indeed some have gone so far as to suggest that it is responsible for more value creation than any other law in U.S. history.³ The breadth of its application has come under increasing fire, however, from critics who contend that it has turned the internet into a lawless realm that is inundated with false and misleading information and often hostile to women and minorities.⁴

* J.D., 2019, New York University School of Law.

** Alfred B. Engelberg Professor of Law and Faculty Director of the Information Law Institute, New York University School of Law. Professor Strandburg is grateful for summer research funding from the Filomen D. Agostino and Max E. Greenberg Research Fund. This Article was prepared for the Symposium entitled *Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law*, hosted by the *Fordham Law Review* and the Neuroscience and Law Center on February 15, 2019, at Fordham University School of Law. For an overview of the Symposium, see Deborah W. Denno & Ryan Surujnath, *Foreword: Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law*, 88 *FORDHAM L. REV.* 381 (2019).

1. 47 U.S.C. § 230 (2012).

2. See, e.g., Derek Khanna, *The Law That Gave Us the Modern Internet—and the Campaign to Kill It*, *ATLANTIC* (Sept. 12, 2013), <https://www.theatlantic.com/business/archive/2013/09/the-law-that-gave-us-the-modern-internet-and-the-campaign-to-kill-it/279588> [<https://perma.cc/3FNB-2UJ3>].

3. David Post, *A Bit of Internet History, or How Two Members of Congress Helped Create a Trillion or so Dollars of Value*, *WASH. POST* (Aug. 27, 2015), <https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/08/27/a-bit-of-internet-history-or-how-two-members-of-congress-helped-create-a-trillion-or-so-dollars-of-value> [<https://perma.cc/HP54-S9M3>].

4. For critiques of CDA 230 and proposals for reform, see, for example, Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revising Section 230 Immunity*, 2 *GEO. L. TECH. REV.* 453 (2018) and Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 *FORDHAM L. REV.* 401, 404–06 (2017) [hereinafter *Internet Will Not Break*] and Benjamin Edelman & Abbey Stemler, *From the Digital to the Physical: Federal Limitations on Regulating Online Marketplaces*, 56 *HARV. J. ON LEGIS.* 142, 142 (2019) (arguing that the CDA “has been stretched beyond recognition to prevent all manner of prudent regulation” and proposing a variety of possible approaches to reform). For an excellent, sympathetic overview of the provision, see Eric Goldman, *An Overview of the United States' Section 230 Internet Immunity*, in *THE OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY* (forthcoming), <https://ssrn.com/abstract=3306737> [<https://perma.cc/UWH7-6PHE>].

This Article analyzes CDA 230 liability in light of the evolution of smart services employing data-driven personalized models of user behavior. As an illustrative case study, we discuss discrimination claims against Facebook's ad-targeting platform, relying on recent empirical studies⁵ and litigation documents for factual background.⁶

We argue that current controversies about CDA 230's scope are best analyzed within a larger context of debates about secondary liability regimes. Viewing CDA 230 through this lens, we propose reforms that account for the growing predominance of smart services and provide pathways for dealing with other areas of concern.

Part I provides background on CDA 230, taxonomizing the primary strains of the case law in a secondary liability frame. Part II describes aspects of Facebook's ad-targeting platform that are salient for our analysis. Part III outlines the basics of discriminatory advertising law under the Fair Housing Act⁷ (FHA) and considers potential liability for two types of audience selection tools. Part IV analyzes CDA 230's applicability to those audience selection tools. Part V analyzes CDA 230 from a secondary liability perspective. This Article concludes with proposals for modifying CDA 230 in light of the preceding analysis.

I. CDA 230 IMMUNITY: HOPES AND STRUGGLES

Section 230(c) of the CDA, entitled "Protection for Good Samaritan Blocking and Screening of Offensive Material," grants providers and users of "interactive computer services" sweeping exemption from liability for actionable content created or published by others.⁸ Its central provision states: "[n]o 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."⁹ This Part briefly reviews CDA 230's history and judicial interpretation and explains how the evolution of the doctrine touches upon themes common to secondary liability regimes.

5. See Muhammad Ali et al., *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes* (Sept. 12, 2019) (unpublished manuscript), <https://arxiv.org/pdf/1904.02095.pdf> [<https://perma.cc/7L47-YLA8>]; see also Amit Datta et al., *Discrimination in Online Advertising: A Multidisciplinary Inquiry*, 81 PROC. MACHINE LEARNING RES. 20 (2018); Till Speicher et al., *Potential for Discrimination in Online Targeted Advertising*, 81 PROC. MACHINE LEARNING RES. 5 (2018).

6. See generally Settlement Agreement and Release, *Nat'l Fair Hous. All. v. Facebook, Inc.*, No. 1:18-cv-02689-JGK (S.D.N.Y. Mar. 28, 2019), ECF No. 67-2 [hereinafter Settlement Agreement]; Statement of Interest of the United States of America, *Nat'l Fair Hous. All.*, No. 1:18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018), ECF No. 48 [hereinafter HUD Brief].

7. Pub. L. No. 90-284, tit. VIII, 82 Stat. 81 (1968) (codified as amended in scattered sections of the U.S.C.).

8. 47 U.S.C. § 230(c) (2012).

9. *Id.*

A. CDA 230's History and Purpose

Congress enacted CDA 230 in 1996, during the early years of the internet's transformation into a hub of commercial and social activity. Its enactment responded to a judicial ruling that Prodigy, which hosted a large set of online "bulletin boards," could be liable as the "publisher" of defamatory user postings.¹⁰ As evidence that Prodigy was a publisher, rather than merely a distributor, of user postings, the ruling relied heavily on Prodigy's employment of content moderators, promulgation of content guidelines, and use of software to filter out offensive language.¹¹

Congress, as explained in the statutory "Findings" and "Policy,"¹² was concerned about the downstream effects of holding service providers liable for users' defamation simply because they attempted to screen out unwelcome content.¹³ Traditional publisher-style screening for actionable content would have been untenable for online services that provided forums for user-driven exchanges involving large amounts of rapidly changing content.¹⁴ Congress also anticipated that, if relieved of liability, online service providers would develop innovative technological "fixes" to the content-screening problem.¹⁵ Given these assumptions, the benefits of publisher liability seemed far outweighed by the costs.

Today, the hope that automated filtering technology could effectively screen user content for actionable defamation, harassment, and the like seems naive. CDA 230's drafters also seem to have vastly underestimated the harm side of the equation and undoubtedly would have been horrified by the tsunami of racist, sexist, homophobic, fraudulent, untruthful, and otherwise hurtful discourse that has accompanied the internet's benefits.

The evolution from hands-off user forums to smart services has led to an interesting jujitsu. While Congress originally feared that online service providers would respond to liability risk by giving up trying to filter out "objectionable" content, proponents of strong CDA 230 immunity now fear that service providers will engage in overly cautious "collateral

10. *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995), *superseded by statute*, Communications Decency Act of 1996, Pub. L. No. 104-104, tit. V, § 230, 110 Stat. 133, 137; *see also* H.R. REP. NO. 104-458, at 194 (1996) (Conf. Rep.) ("One of the specific purposes of this section is to overrule *Stratton-Oakmont v. Prodigy* and any other similar decisions which have treated such providers and users as publishers or speakers of content that is not their own because they have restricted access to objectionable material.").

11. *Stratton*, 1995 WL 323710, at *5.

12. 47 U.S.C. § 230(a)-(b) (2012).

13. H.R. REP. NO. 104-458, at 86.

14. *Internet Will Not Break*, *supra* note 4, at 404-06 (discussing the background to CDA 230 and noting Prodigy's argument that "it could not possibly edit the thousands of daily messages posted to its bulletin boards as a traditional publisher would").

15. 47 U.S.C. § 230(b) (listing policy objectives including "to encourage the development of technologies which maximize user control over what information is received" and "to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material").

ensorship.”¹⁶ Bolstering this fear, large online service providers have responded to calls for regulation by constructing extensive regimens combining filtering with human review, which raises freedom of expression concerns.¹⁷

B. Judicial Interpretation and CDA 230’s Broad Sweep

CDA 230 immunizes any defendant who is “treated as the publisher or speaker of any information provided by another information content provider.”¹⁸ Because CDA 230 does not define “publisher,” its interpretation has been a central, and difficult, task for the courts. Plaintiffs initially responded to CDA 230 defenses by arguing that the provision preserves defamation doctrine’s distinction between “publishers,” who are held responsible if they negligently publish defamatory content, and “distributors,” who are liable only if they receive notice and do not act.¹⁹ Early cases rejected this distinction, however, immunizing service providers against both “publisher” and “distributor” liability²⁰ and giving them free rein in handling requests to take down allegedly actionable content.²¹

These early cases were harbingers of broad judicial interpretation of “publisher” under CDA 230. Many cases have tested the scope of “publisher” activities, with results holding, for example, that CDA 230 immunizes decisions about what to post; nonsubstantive editing; reformatting of fonts, colors, and the like; and re-presentation of information in the form of star ratings or maps.²²

16. See generally Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293 (2011); James Grimmelman, *To Err Is Platform*, KNIGHT FIRST AMEND. INST. COLUM. U. (Apr. 6, 2018), <https://knightcolumbia.org/content/err-platform> [<https://perma.cc/52C7-TJGL>].

17. See Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1635 (2018).

18. 47 U.S.C. § 230(c)(1) (2012). The reference to “speakers” is essentially redundant and has played little role in CDA 230’s interpretation.

19. See, e.g., *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 332 (4th Cir. 1997); *Blumenthal v. Drudge*, 992 F. Supp. 44, 49–50 (D.D.C. 1998).

20. See, e.g., *Zeran*, 129 F.3d at 334; *Blumenthal*, 992 F. Supp. at 52.

21. For a particularly expansive, and controversial, ruling that CDA 230 applies even to takedown injunctions issued after users have been held liable, see generally *Hassell v. Bird*, 420 P.3d 776 (Cal. 2018).

22. See, e.g., *Marshall’s Locksmith Serv. Inc. v. Google, LLC*, 925 F.3d 1263, 1269–71 (D.C. Cir. 2019) (finding CDA immunity even where Google put the advertisements into a map format); *Kimzey v. Yelp! Inc.*, 836 F.3d 1263, 1269–70 (9th Cir. 2016) (finding CDA immunity even where Yelp! took reviews from a different website and added a star rating); *O’Kroley v. Fastcase, Inc.*, 831 F.3d 352, 355 (6th Cir. 2016) (finding CDA immunity even where Google had performed some “automated editorial acts on the content, such as removing spaces and altering font” and “kept the search result up even after [the plaintiff] complained about it”); *Jones v. Dirty World Entm’t Recordings LLC*, 755 F.3d 398, 416 (6th Cir. 2014) (“The CDA expressly bars ‘lawsuits seeking to hold a service provider liable for its exercise of a publisher’s traditional editorial functions such as deciding whether to publish, withdraw, postpone or alter content.’” (quoting *Zeran*, 129 F.3d at 330)).

Courts also have interpreted CDA 230's substantive range of applicability broadly,²³ except in a few arenas where it was explicitly limited.²⁴ While scholars and policymakers have proposed removing CDA 230 protections from additional substantive categories,²⁵ only the recent Allow States and Victims to Fight Online Sex Trafficking Act of 2017²⁶ (FOSTA) has been enacted thus far.

Though refusing to recognize broad substantive exceptions to CDA 230's coverage, courts have sometimes declined to extend CDA 230 protection because particular claims did not treat the defendant as a "publisher." For example, courts have denied immunity from claims of failure to provide warnings required by state law,²⁷ violations of federal robocalling regulations,²⁸ failure to follow regulations applicable to rental agents,²⁹ failure to collect city amusement tax,³⁰ and failure to post ads after accepting payment.³¹ Consistent with this approach, courts have emphasized that CDA 230 does not provide blanket protection to a service provider who "plays

23. *See, e.g.*, *Herrick v. Grindr LLC*, 765 F. App'x 586 (2d Cir. 2019) (state torts including negligence, deceptive business practices and false advertising, intentional and negligent infliction of emotional distress, failure to warn, and negligent misrepresentation); *Chi. Lawyers' Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666 (7th Cir. 2008) (Fair Housing Act); *Gonzalez v. Google, Inc.*, 335 F. Supp. 3d 1156 (N.D. Cal. 2018) ("material support" for terrorism); *Noah v. AOL Time Warner Inc.*, 261 F. Supp. 2d 532 (E.D. Va. 2003) (Title II of the Civil Rights Act); *Hassell*, 420 P.3d 776 (enforcement of a takedown injunction); *Hill v. StubHub, Inc.*, 727 S.E.2d 550 (N.C. 2012) (unfair or deceptive trade practices).

24. 47 U.S.C. § 230(e) (2012) (limiting immunity for claims involving federal criminal law, intellectual property law, and electronic surveillance law).

25. *See, e.g.*, Nancy Leong & Aaron Belzer, *The New Public Accommodations: Race Discrimination in the Platform Economy*, 105 GEO. L.J. 1271, 1320–21 (2017); Press Release, Nat'l Ass'n of Attorneys Gen., NAAG Supports Amendment to the Communications Decency Act (May 23, 2019), <https://www.naag.org/naag/media/naag-news/naag-supports-amendment-to-the-communications-decency-act.php> [<https://perma.cc/728T-PLB6>] (suggesting an exception for state criminal violations); Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMEND. INST. COLUM. U. (Apr. 1, 2018), <https://knightcolumbia.org/content/discriminatory-designs-user-data> [<https://perma.cc/8H5G-732C>] (suggesting "an explicit exception to Section 230 immunity for violations of civil rights laws").

26. Pub. L. No. 115-164, 132 Stat. 1253 (codified in scattered sections of 18 and 47 U.S.C.).

27. *See, e.g.*, *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 850–52 (9th Cir. 2016) (denying immunity in a failure to warn case where particular third parties used the defendant's website to target and lure rape victims).

28. *See, e.g.*, *Cunningham v. Montes*, 378 F. Supp. 3d 741, 750 (W.D. Wis. 2019) (denying immunity in a Telephone Consumer Protection Act claim based on robocalling).

29. *See, e.g.*, *HomeAway.com, Inc. v. City of Santa Monica*, 918 F.3d 676, 682–84 (9th Cir. 2019) (denying immunity in a case involving rental regulations).

30. *See, e.g.*, *City of Chicago v. StubHub!, Inc.*, 624 F.3d 363, 366 (7th Cir. 2010) (noting that "Chicago's amusement tax does not depend on who 'publishes' any information or is a 'speaker'" and thus "Section 230(c) is irrelevant").

31. *See, e.g.*, *O'Hara-Harmon v. Facebook, Inc.*, No. 19-cv-00601 WHA, 2019 WL 1994087, at *3 (N.D. Cal. May 5, 2019) ("Facebook would likely not be immunized from a claim alleging it charged and collected money to publish advertising that it then did not publish.").

multiple related but distinct roles in the online platform it has established.”³² For example, where Airbnb acted as both “a publisher of third-party rental listings” and “an agent that books rental agreements between users and hosts and collects and distributes payments,” it was not immune from regulations “directed only at Airbnb’s conduct in the latter role.”³³

Faced with the broad judicial interpretation of “publisher,” plaintiffs often contend that a service provider is liable because it acted as an “information content provider” in its own right. CDA 230 defines “information content provider” as “any 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.”³⁴ Courts have routinely treated service providers as potentially liable “information content providers” when they or their employees independently created or developed actionable content.³⁵ At the other end of the spectrum, courts have generally granted CDA 230 protection to defendants when plaintiffs alleged merely that a provider knew that a service was being used for illegal purposes or profited from a third party’s creation and publication of actionable content.³⁶

In intermediate cases, the crux of the issue is whether various activities amount to “development” of information content “in part.”³⁷ Though CDA 230 doctrine has not drawn a bright line, it is useful to consider two lines of argument which, foreshadowing our analysis in Part V, we can analogize to separate threads of secondary liability.³⁸ One argument, analogous to “contributory liability,” contends that the design or structure of a service contributes to the information content’s alleged illegality, thereby making the provider partly responsible for its development.³⁹ A second argument, analogous to “inducement,” contends that a provider is partly responsible for development when it actively encourages users to develop actionable content.⁴⁰ Courts have generally been unconvinced by inducement-type

32. *HomeAway.com*, 918 F.3d at 683–84; *Airbnb, Inc. v. City of Boston*, No. 18-12358-LTS, 2019 U.S. Dist. LEXIS 74823, at *11–12 (D. Mass. May 3, 2019), *appeal filed*, No. 19-1561 (1st Cir. June 6, 2019); *see also* *Airbnb, Inc. v. City and County of San Francisco*, 217 F. Supp. 3d 1066, 1072–73 (N.D. Cal. 2016).

33. *Airbnb*, 2019 U.S. Dist. LEXIS 74823, at *11–12.

34. 47 U.S.C. § 230(f)(3) (2012).

35. *See, e.g.*, *Huon v. Denton*, 841 F.3d 733, 742 (7th Cir. 2016); *FTC v. LeadClick Media, LLC*, 838 F.3d 158, 173 (2d Cir. 2016); *FTC v. Accusearch Inc.*, 570 F.3d 1187 (10th Cir. 2009).

36. *See, e.g.*, *Universal Commc’n Sys., Inc. v. Lycos, Inc.*, 478 F.3d 413, 421 (1st Cir. 2007); *Daniel v. Armslist, LLC*, 926 N.W.2d 710, 727 (Wis. 2019).

37. *See* *Fair Hous. Council v. Roommates.com, LLC*, 521 F.3d 1157, 1166–68 (9th Cir. 2008) (en banc).

38. *See infra* Part V; *see also* *Chi. Lawyers’ Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666, 670 (7th Cir. 2008) (relating CDA 230 to contributory copyright infringement).

39. *See* *Craigslist*, 519 F.3d at 670.

40. *See* *Roommates.com*, 521 F.3d at 1165 (“The CDA does not grant immunity for inducing third parties to express illegal preferences.”); *see also* *Lycos*, 478 F.3d at 421 (leaving open the question of whether CDA 230 would immunize the “clear expression or other affirmative steps taken to foster’ unlawful activity that would be necessary to find active inducement” (quoting *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919 (2005))).

arguments. Disagreement about the implications of inducement seems to underlie disparate decisions about whether Backpage.com could rely on CDA 230 immunity with respect to claims involving its hosting of ads for prostitution.⁴¹

Courts have more often held providers responsible for developing actionable information based on the designs of their services. The Ninth Circuit’s en banc decision in *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*⁴² is the seminal case in this regard. That decision held that CDA 230 did not exempt Roommates.com from FHA liability when its website “ma[de] answering the discriminatory questions a condition of doing business” and its search system was designed to “steer users based on the preferences and personal characteristics that Roommate.com itself forces subscribers to disclose,” thus making it “more difficult or impossible for individuals with certain protected characteristics to find housing—something the law prohibits.”⁴³ Roommates.com was, however, immune from liability for allegations concerning posts that users entered into a box designed to collect open-ended comments.⁴⁴

The majority contrasted services designed as *neutral tools* that can be used for both illegal and legal purposes with services that make *material contributions* to the illegality.⁴⁵ Thus, a standard search engine is a neutral tool protected by CDA 230 even if its algorithm sometimes puts a defamatory post at the top of search results.⁴⁶ It is also not enough, under this test, to show that a service *could have been designed* to be less amenable to actionable use.⁴⁷

As of this writing, we are unaware of any judicial opinion interpreting how CDA 230 applies when “smart” algorithms play a role in the illegality associated with online information content. The Department of Housing and Urban Development (HUD), however, has expressed its view in litigation that CDA 230 does not immunize Facebook from liability for discriminatory housing ad targeting, a question we analyze in Part IV.⁴⁸

41. See *J. S. v. Vill. Voice Media Holdings, L.L.C.*, 359 P.3d 714, 718 (Wash. 2015) (finding that CDA immunity would not apply where Backpage.com’s “content requirements are specifically designed to control the nature and context of those advertisements so that pimps can continue to use Backpage.com to traffic in sex, including the trafficking of children”); see also *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 24 (1st Cir. 2016) (finding that CDA immunity did apply to Backpage.com). Similar claims would now be viable under FOSTA.

42. 521 F.3d 1157 (9th Cir. 2008) (en banc).

43. *Id.* at 1166–67.

44. *Id.* at 1173–74.

45. *Id.* at 1167–68.

46. See *O’Kroley v. Fastcase, Inc.*, 831 F.3d 352, 355 (6th Cir. 2016).

47. See *Daniel v. Armslist, LLC*, 926 N.W.2d 710, 722 (Wis. 2019).

48. HUD Brief, *supra* note 6, at 14–17.

II. FACEBOOK'S AD-TARGETING PLATFORM

A. Overview

Facebook supports itself by selling advertising that is targeted using the vast amounts of personal information it gleans from users' account information, friend networks, posts and activities on the Facebook site itself, information about other online activities that Facebook obtains about some users through various tracking mechanisms, and offline data sources.⁴⁹ Facebook sells advertising through a sophisticated online platform that can be used by advertisers, large and small, to target ads to prospective customers.⁵⁰ Our description is a snapshot intended as a springboard for our analysis, rather than a representation of the current status of Facebook's ad platform, which is modified often. It is based on the recent empirical work of Muhammad Ali and Till Speicher.⁵¹ Before and subsequent to these empirical studies, for example, Facebook changed the platform in response to press reports and claims made in litigation.⁵² More major modifications are in the works as part of a 2019 settlement agreement and will be discussed below.⁵³

The process by which a particular ad is placed before a particular user has a number of steps.⁵⁴ First, advertisers supply the content for the ads, choose from among several approaches for selecting a target audience, and provide information about their advertising budgets and goals.⁵⁵ Second, ads are eventually placed using an auction algorithm which considers advertisers' objectives and advertising budgets,⁵⁶ along with "estimated action rates" and "ad quality and relevance."⁵⁷ Facebook's algorithms for ad targeting and placement are, of course, proprietary.⁵⁸

B. Audience Selection Approaches

Facebook offers advertisers several approaches to audience selection. Here we focus on its two "smart" approaches: attribute-based targeting and

49. See generally Ali et al., *supra* note 5.

50. *Id.*

51. See *id.*; see also Datta et al., *supra* note 5; Speicher et al., *supra* note 5; Julia Angwin et al., *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, PROPUBLICA (Nov. 21, 2017, 1:23 PM), <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin> [<https://perma.cc/UY3V-4923>].

52. See, e.g., Sheryl Sandberg, *Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising*, FACEBOOK NEWSROOM (Mar. 19, 2019), <https://newsroom.fb.com/news/2019/03/protecting-against-discrimination-in-ads> [<https://perma.cc/9P3A-9KHD>].

53. See *infra* Part II.B.4.

54. Ali et al., *supra* note 5.

55. *Id.* at 4.

56. *Id.*

57. *Id.*

58. See generally *id.*

“lookalike audience” targeting. Empirical studies demonstrate that each can, at least in principle, result in biased targeting.⁵⁹

1. Attribute-Based Targeting

Facebook’s attribute-based audience selection permits advertisers to select audiences based on thousands of attributes which Facebook creates based on information gleaned from its users’ activities. While some of these “attributes” are based straightforwardly on Facebook users’ self-identification in their “profiles,” most are defined and constructed by Facebook based on data-driven analysis of users’ activities.⁶⁰ Facebook offers a menu of attributes divided into categories: demographics, interests, and behaviors.⁶¹ Users can also search for and employ hundreds of thousands of additional “free-form” attributes.⁶² As of the time of this writing, the demographic menu, for example, includes categories such as education, finances, life events, parents, relationships, and work, which can be drilled down to more specific attributes such as education level or schools attended.⁶³ Most of these attributes can be used both to include and to exclude users.⁶⁴

The menu for behavioral attributes includes a very large number of categories, perhaps the most interesting of which is “multicultural affinity.” According to Facebook, this category is not based on a user’s ethnic identity but, rather, “represents how interested the user is in content related to different ethnic communities.”⁶⁵ Originally called “ethnic affinity,” it was renamed and removed from the list of possible exclusions after a ProPublica article exposed that “ethnic affinity” could be used to exclude particular ethnic groups from targeted audiences.⁶⁶ Options for “multicultural affinity” include “African-American, Asian-American, Hispanic-All, Hispanic-English Dominant, Hispanic-Spanish Dominant, [and] Hispanic-Bilingual.”⁶⁷

While certain attributes, such as multicultural affinity, age, gender, or religion, are either protected characteristics or obviously close proxies, there

59. See generally Datta et al., *supra* note 5; Speicher et al., *supra* note 5; Ali et al., *supra* note 5; Angwin et al., *supra* note 51.

60. Speicher et al., *supra* note 5, at 11 (“It is unclear how exactly Facebook infers these attributes, but from their own description this information can be gathered in many different ways such as user activity on Facebook pages, apps and services, check-ins with Facebook, and accesses to external webpages that use Facebook ad technologies.”).

61. *Id.* at 7.

62. *Id.* at 11–12.

63. See generally *id.*; *Help Your Ads Find the People Who Will Love Your Business*, FACEBOOK BUS., <https://www.facebook.com/business/ads/ad-targeting> [<https://perma.cc/QYT5-69TX>] (last visited Oct. 6, 2019).

64. See Speicher et al., *supra* note 5, at 7.

65. *Id.* at 6.

66. *Id.*

67. Jeremy B. Merrill, *What We Learned from Collecting 100,000 Targeted Facebook Ads*, PROPUBLICA: ELECTIONLAND (Dec. 26, 2018, 12:39 PM), <https://www.propublica.org/article/facebook-political-ad-collector-targeted-ads-what-we-learned> [<https://perma.cc/S8YC-Y2MZ>].

are also less obvious attributes or combinations of attributes that function as reasonably close proxies for such protected characteristics.⁶⁸

2. Lookalike Audience

To use the lookalike audience approach, an advertiser provides Facebook with a base audience list and Facebook's algorithms create an expanded audience of "similar" people.⁶⁹ While it is not obvious that demographic similarity is generally relevant to creating promising audiences for targeted ads, Speicher has demonstrated that Facebook's lookalike audiences tend to reproduce demographic disparity in the base audience.⁷⁰

3. Changes Required by the 2019 Consent Decree

A class action lawsuit filed by the National Fair Housing Alliance (NFHA) against Facebook settled with a consent decree issued in March 2019.⁷¹ Among other changes, the consent decree requires Facebook to limit the ways in which the tools outlined above can be used for targeting housing, employment, or credit (HEC) ads.⁷² On or before September 30, 2019, Facebook will implement a tool called HEC Flow, which will modify the audience selection tools in four ways.⁷³ First, targeting by zip code will no longer be an option. Instead, geographic targeting will have to encompass at least a fifteen-mile radius from the center of a chosen city or address.⁷⁴ Second, age and gender will no longer be available as targeting criteria for those categories of ads.⁷⁵ Third, Facebook promises that "[n]o targeting options that Facebook determines are direct descriptors of, or semantically or conceptually related to, a person or group of people based on Protected Classes will be available" for targeting HEC ads.⁷⁶ The consent decree defines "direct descriptors" as "targeting options whose names directly describe persons in Protected Classes" and defines "semantically or conceptually related to" as "targeting options whose names appear to be associated with a Protected [Class]."⁷⁷

Fourth, the consent decree specifies that the lookalike audiences algorithm can use data from the country, region, profession, and field of study of a user

68. For example, Speicher found that facially neutral attributes, such as "U.S. Politics: Very Liberal" and "Interest: Online games," were disproportionately correlated with Facebook profiles belonging to Black Facebook users, while facially neutral attributes such as "U.S. Politics: Very Conservative" and "Interest: Hiking" were disproportionately correlated with Facebook profiles belonging to White Facebook users. Speicher et al., *supra* note 5, at 12.

69. *Id.* at 15.

70. *Id.* at 15–18.

71. Settlement Agreement, *supra* note 6, at 1.

72. *Id.* at 3–4.

73. *Id.* at 4.

74. *Id.*

75. *Id.*

76. *Id.* Exhibit A, at 3.

77. *Id.*

profile but may not consider “age, gender, relationship status, religious views, school, political views, interested in, or zip code.”⁷⁸ The lookalike audiences tool may, however, continue to use all data about user behavior on the site (such as ad engagement, apps, or pages), with the exception of Facebook groups.⁷⁹ The decree further requires Facebook to rename the tool so that it does not “refer to finding users who ‘look like’ users provided by advertisers.”⁸⁰ It also requires that Facebook provide a “Housing Search Portal,” such that users will be able to search all housing ads posted on Facebook, regardless of whether they are in the targeted audience.⁸¹

III. HOUSING DISCRIMINATION LIABILITY AND AUTOMATED AD-TARGETING PLATFORMS

In some contexts, such as housing, employment, and credit, U.S. laws prohibit private party discrimination, including discriminatory advertising, based on “protected” attributes, which commonly include race, color, religion, national origin, sex, age, and disability. Our case study focuses on discriminatory housing advertising, which is prohibited under the Fair Housing Act. Part III.A explains the FHA’s standards for determining whether actionable discriminatory advertising has occurred and who can be held liable for actionable discrimination, highlighting unsettled questions about the standard that would apply to disparate ad targeting. Part III.B then considers whether and how an ad-targeting platform such as Facebook might be held liable for discrimination resulting from attribute-based or lookalike audience ad targeting. In this Part, we set aside the question of CDA 230 protection, which is addressed in Part V.

A. Discriminatory Ad Targeting Under the Fair Housing Act

The FHA makes it unlawful:

to make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin, or an intention to make any such preference, limitation, or discrimination.⁸²

We divide our discussion of potential liability into two inquiries. First, has actionable discrimination occurred under the FHA? Second, if so, who is liable?

1. Has Actionable Discrimination Occurred?

Discriminatory housing advertising claims are generally evaluated based on whether an “ordinary reader” of the ad would have perceived an unlawful

78. *Id.* at 4.

79. *Id.*

80. *Id.*

81. *Id.*

82. 42 U.S.C. § 3604(c) (2012).

“preference, limitation, or discrimination.”⁸³ The FHA’s “ordinary reader” standard stands out from other discrimination standards under the FHA and elsewhere because the defendant’s intentions, purposes, or reasons are largely irrelevant under the standard, which has been characterized as a “strict liability” provision.⁸⁴ Although most of the case law under these provisions involves advertising *content*, discriminatory ad *targeting* is also actionable according to regulations promulgated by HUD, which is responsible for enforcing the FHA.

Under the regulations, discriminatory advertising includes “[s]electing media or locations for advertising the sale or rental of dwellings which deny particular segments of the housing market information about housing opportunities because of race, color, religion, sex, handicap, familial status, or national origin.”⁸⁵ The regulations also prohibit “limiting information, by word or conduct, regarding suitably priced dwellings available for inspection, sale or rental, because of race, color, religion, sex, handicap, familial status, or national origin.”⁸⁶

Former HUD guidelines provide the following further examples of actionable selective advertising:

[T]he use of English language media alone or the exclusive use of media catering to the majority population in an area, when, in such area, there are also available non-English language or other minority media, may have discriminatory impact. . . . The following are examples of the selective use of advertisements which may be discriminatory:

(a) *Selective geographic advertisements.* Such selective use may involve the strategic placement of billboards; brochure advertisements distributed within a limited geographic area by hand or in the mail; advertising in particular geographic coverage editions of major metropolitan newspapers or in newspapers of limited circulation which are mainly advertising vehicles for reaching a particular segment of the community; or displays or announcements available only in selected sales offices.⁸⁷

Since the effects of discriminatory targeting are the same whether it is accomplished by selective use of particular media or by automated targeting, it is unsurprising that HUD has taken the position that the FHA’s discriminatory advertising prohibitions reach such automated targeting.⁸⁸

83. See, e.g., *Ragin v. Harry Macklowe Real Estate Co.*, 6 F.3d 898, 905 (2d Cir. 1993) (*Ragin II*); *Ragin v. N.Y. Times Co.*, 923 F.2d 995, 999–1000 (2d Cir. 1991) (*Ragin I*).

84. See, e.g., Stephanie M. Stern, *A Social Norm Theory of Regulating Housing Speech Under the Fair Housing Act*, 84 MO. L. REV. 435, 441–45 (2019).

85. 24 C.F.R. § 100.75(c)(3) (2019).

86. 24 C.F.R. § 100.80(b)(4) (2019).

87. 24 C.F.R. § 109.25 (1995) (withdrawn from the C.F.R. by Office of the Assistant Secretary for Fair Housing and Equal Opportunity; Regulatory Reinvention; Streamlining of HUD’s Regulations Implementing the Fair Housing Act, 61 Fed. Reg. 14,378 (Apr. 1, 1996) (to be codified at 24 C.F.R. pts. 100, 103, and 109)).

88. See HUD Brief, *supra* note 6, at 23–24; *HUD v. Facebook, Inc.*, FHEO No. 01-18-0323-8, at 1–2 (Charge of Discrimination Mar. 28, 2019), <https://www.hud.gov/>

There are, as yet, no judicial opinions that apply the FHA to automated behavioral ad targeting and it remains unclear what standard courts should apply to assess whether disparate ad targeting is discriminatory under the FHA. HUD invoked the ordinary reader standard in a recent brief in the Facebook advertising litigation, but neither that brief nor the few judicial opinions touching on more traditional ad targeting provide much explanation of how the ordinary reader standard should be applied.⁸⁹

Traditionally, the ordinary reader standard turns on readers' likely perceptions of the content of the ad, regardless of advertisers' intentions.⁹⁰ Discriminatory ad targeting cannot be assessed on the basis of ad content, however, which begs the question of what the ordinary reader would look at to judge whether actionable discrimination has occurred. One possible approach would consider whether the demographic makeup of the ultimate target audience would indicate a prohibited "preference, limitation, or discrimination" to an ordinary reader.⁹¹ This first approach seems in tune with the ordinary reader standard's broad remedial approach but raises questions about whether an ordinary reader has a reasonable basis for judging ad audience demographics. An alternative approach might instead consider the ordinary reader's perspective on the *ad-targeting plan* or process.⁹² This second approach also raises questions about the competence of ordinary readers, particularly where complicated automated targeting tools are used. In addition, it might belie the purpose of the ordinary reader standard by lacking sensitivity to the targeting's real-world impact on potential customers.

Given that discriminatory ad targeting may have little or nothing to do with how readers perceive ad content, courts might instead set aside the ordinary reader approach in favor of the disparate treatment or impact approaches applied to other aspects of housing discrimination.⁹³ Under these theories, a burden-shifting approach is applied. Plaintiffs must first prove a *prima facie* case showing either that they were treated differently based on a protected characteristic or that a practice (here, ad targeting) disparately impacted a protected class.⁹⁴ The defendant may then either rebut the *prima facie* case

sites/dfiles/FHEO/documents/HUD%20v%20Facebook%20-%20Charge.pdf
[<https://perma.cc/BT8Q-M6RX>] [hereinafter HUD Charge].

89. See HUD Brief, *supra* note 6, at 10–11.

90. See, e.g., *Miami Valley Fair Hous. Ctr., Inc. v. Connor Grp.*, 805 F. Supp. 2d 396, 407 (S.D. Ohio 2011).

91. *Id.* at 406 (quoting Fair Housing Act, 42 U.S.C. § 3604(c) (2006)).

92. *Id.* at 411 ("In cases like this, where the ads in question are part of a large advertising campaign, but are not facially discriminatory, courts have found it appropriate to look at the entire advertising campaign to help determine whether there is an FHA violation.")

93. See *Tex. Dep't of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507, 2518–26 (2015).

94. *Id.* at 2523. For an overview of these approaches and a discussion of how they might apply to data-driven models, see Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 694–712 (2016). Note that HUD has recently issued a proposed rule that would heighten the standard under 24 C.F.R. § 100.500 for a *prima facie* case based on disparate impact. HUD's Implementation of the Fair Housing Act's Disparate Impact Standard, 84 Fed. Reg. 42,854, 42,858–59 (Aug. 19, 2019) (to be codified at 24 C.F.R.

or produce evidence that there was a nondiscriminatory “legitimate reason” or “business necessity” for the treatment or practice.⁹⁵ Plaintiffs may rebut the proffered reason by demonstrating that it is pretextual or, in a disparate impact analysis, that there is a reasonable, less discriminatory alternative.⁹⁶ The outcomes of particular cases will, of course, turn upon which of these (or some other) standards courts eventually adopt.

2. Who Is Liable?

24 C.F.R. § 100.7 governs who can be held liable, assuming actionable discrimination can be established.⁹⁷ Direct liability is broadly construed to include not only a person’s “own conduct that results in a discriminatory housing practice” but also a person’s “fail[ure] to take prompt action to correct and end a discriminatory housing practice by a third-party, where the person knew or should have known of the discriminatory conduct and had the power to correct it.”⁹⁸

Courts have held not only advertisers but also intermediaries, such as newspapers and listing services, liable for discriminatory advertising.⁹⁹ A recent opinion considered the potential liability of a tenant-screening service

pt. 100). The proposed rule also provides standards for defendants to rebut the plaintiff’s prima facie case if it alleges that the “cause of a discriminatory effect is a model used by the defendant, such as a risk assessment algorithm.” *Id.* at 42,862. If adopted as final, the proposed rule changes are expected to make it significantly more difficult for plaintiffs to prove discrimination under a disparate impact theory and to restrict the claims that can be brought based on defendants’ use of automated algorithms. *See, e.g.,* Emily Badger, *Who’s to Blame When Algorithms Discriminate?*, N.Y. TIMES (Aug. 20, 2019), <https://www.nytimes.com/2019/08/20/upshot/housing-discrimination-algorithms-hud.html> [https://perma.cc/T23L-JE8P]; Lola Fadulu, *Trump Proposal Would Raise Bar for Proving Housing Discrimination*, N.Y. TIMES (Aug. 2, 2019), <https://www.nytimes.com/2019/08/02/us/politics/trump-housing-discrimination.html> [https://perma.cc/XR3K-ENK6]. It appears, however, that the proposed rule changes do not have substantial implications for the arguments made in this Article, which mostly do not depend on specifics of the disparate impact standard. In particular, the proposed rule changes do not amend either 24 C.F.R. § 100.75, which deals with discriminatory advertising, or any of the relevant provisions of 24 C.F.R. § 100.7, which determines who can be held liable for discrimination under the FHA. It is also worth noting that the HUD charge against Facebook, which was issued not long before the proposed rule, claims discrimination based on both attribute-based and lookalike audience targeting, which suggests that HUD itself does not view its proposed rule changes as inconsistent with ad-targeting platform liability under the FHA.

95. *See Inclusive Cmty. Project*, 135 S. Ct. at 2522–23; *see also* Barocas & Selbst, *supra* note 94, at 696, 701.

96. *See Inclusive Cmty. Project*, 135 S. Ct. at 2518.

97. The regulation was adopted in 2016, but, according to HUD’s responses to comments, “Section 100.7 does not create liability that does not already exist” and “helps clarify the Act’s coverage for residents and housing providers.” *Quid Pro Quo and Hostile Environment Harassment and Liability for Discriminatory Housing Practices Under the Fair Housing Act*, 81 Fed. Reg. 63,054, 63,066–67 (Sept. 14, 2016) (to be codified at 24 C.F.R. pt. 100). It provides for both direct liability in section (a) and vicarious liability for the actions of an agent or employee in section (b). Here we concern ourselves only with direct liability.

98. 24 C.F.R. § 100.7 (2019).

99. *See, e.g., Ragin I*, 923 F.2d 995, 1000 (2d Cir. 1991).

that provided automated criminal records analyses to landlords.¹⁰⁰ While not based on discriminatory *advertising*, this opinion is instructive regarding the question of who is directly liable under the FHA.¹⁰¹ The court held that the screening service's own conduct was actionable because "[a]llowing a screening company to facilitate discrimination by disqualifying qualified applicants on an impermissible basis or by allowing a customer to set impermissible qualification standards with impunity would subvert the purpose of the FHA."¹⁰² Alternatively, the service could be liable for failing to correct the landlord's discriminatory conduct because it had "held itself out as a company with the knowledge and ingenuity to screen housing applicants by interpreting criminal records and specifically advertised its ability to improve 'Fair Housing compliance'" and had "a duty not to sell a product to a customer which would unwittingly cause its customer to violate federal housing law and regulations."¹⁰³

B. *Ad-Targeting Platform Liability Under the FHA*

This section explores how the FHA's liability standards might apply to disparate ad targeting resulting from the use of attribute-based and lookalike audience ad-targeting tools. As discussed in Part III.A, the FHA standard for discriminatory ad targeting is unsettled and certainly raises interesting questions. Here, however, our exploration of FHA ad-targeting liability is aimed at illuminating our understanding of how CDA 230's breadth affects substantive regulation, particularly for "smart" services. Because CDA 230 only matters when service providers would otherwise be held liable, the discussion generally assumes actionable discrimination has occurred and focuses on whether ad-targeting platforms are likely to be liable for that discrimination. It delves into the potential substantive standards only to the extent that they have implications for that question.

This section first considers the possibility that smart platforms such as Facebook generally have so much information about how ads are targeted to their users that they should essentially always be held liable for failure to correct advertisers' discriminatory ad targeting, regardless of the targeting process. Setting that possibility aside, it then considers the potential for ad-platform liability under the FHA based on advertisers' use of attribute-based and lookalike audience selection tools.

1. Should Smart Platforms Always Be Liable for Failure to Correct?

As noted above, FHA liability can attach not only for a defendant's own conduct but also for failure to correct discriminatory conduct that the defendant knew or should have known about and had the power to correct.¹⁰⁴

100. *See generally* Conn. Fair Hous. Ctr. v. Corelogic Rental Prop. Sols., LLC, 369 F. Supp. 3d 362 (D. Conn. 2019).

101. *Id.*

102. *Id.* at 372.

103. *Id.*

104. *See supra* Part III.A.

Ad-targeting platforms have complete power over the design of their ad-placement systems and the eventual placement of ads.¹⁰⁵ In this respect, they are much like the tenant-screening service held responsible when its automated tool produced criminal history evaluations used in a landlord's discriminatory decisions.¹⁰⁶ Arguably, a platform's ultimate control over its ad placement system always gives it sufficient "power to correct" discriminatory ad targeting,¹⁰⁷ even though it does not control the various choices its advertisers make when using its audience selection tools. If so, an ad-targeting platform's liability under the FHA will turn on whether it "should have known" about an advertiser's discriminatory practices.¹⁰⁸

A smart ad-targeting platform, such as Facebook, has extensive information about how ads are targeted on its platform: it knows what attributes its advertisers select, can analyze ad content, and has a wealth of relevant knowledge about its users, both within and outside of the audience that is ultimately targeted.¹⁰⁹ Arguably, this data and technical capacity mean that a smart ad-targeting platform "should know" whenever any targeted audience has a significantly disparate makeup.¹¹⁰ Liability would then flow from the platform's failure to correct the disparity by readjusting the audience. Though this argument for platform liability is perhaps novel in the FHA context, it is not as extreme as it might sound. A smart platform is often the *only* party with either the capacity to detect discriminatory ad targeting or the power to correct it. Moreover, this understanding of the "failure to correct" standard would amount, in essence, to imposing a duty on ad-targeting platforms that is rather similar to the negligence standard applied to traditional publishers.

Of course, courts may not adopt such a sweeping view of ad-targeting platform power and knowledge. Even so, "failure to correct" liability might arise from the features of particular targeting tools, as discussed during the evaluation of liability for attribute-based and lookalike audience targeting in the next two subsections.

2. FHA Liability Arising from Attribute-Based Targeting

While we are primarily interested in platform liability, we begin by discussing advertiser liability, as a point of comparison and a potential basis for a platform's liability for "failure to correct."

105. *See supra* Part II.

106. *See Corelogic*, 369 F. Supp. 3d at 374.

107. 24 C.F.R. § 100.7(a)(1)(iii) (2019).

108. *Id.*

109. *See supra* Part II.B.

110. How to determine what a smart platform "should know" is a fascinating question, which we do not consider further in this Article.

a. Advertiser Liability

Courts would almost certainly hold housing advertisers liable for choosing targeting attributes that are direct or reasonably apparent proxies for protected characteristics, either under an ordinary reader standard or because advertisers would be hard-pressed to come up with legitimate reasons for such targeting. If disparate targeting results from less obviously problematic attribute choices, however, the outcome might depend on the substantive standard. We can presume for sake of argument that the disparity in the ultimate target audience would indicate a “preference, limitation, or discrimination” to an “ordinary reader” or create a prima facie case of disparate impact.¹¹¹ Moreover, a housing advertiser might have considerable difficulty providing legitimate reasons for targeting based on attributes that bear no apparent relationship to protected characteristics if they also seem unrelated to an individual’s likely interest in a particular housing ad or suitability as a tenant or purchaser. Many, if not most, of the attributes that Facebook makes available for ad targeting are of this ilk. Under a standard based on the ordinary reader’s perception of an advertiser’s *targeting plan*, however, plaintiffs might be much less successful.

b. Platform Liability

For attribute-based targeting, the platform’s “own conduct” entails choosing, defining, and computing the attributes its tool offers and making them available to housing advertisers.¹¹² In a “but-for” sense, this conduct “results in” any actionable targeting that occurs, but such a sweeping interpretation is probably beyond the scope of even the FHA’s “strict liability” approach to advertising discrimination. Distinctions can be drawn between various types of ad-targeting platform conduct relating to attribute-based targeting.

111. Making the prima facie case would be more difficult under HUD’s proposed modifications to 24 C.F.R. § 100.500, particularly because it requires plaintiffs to state facts plausibly alleging that a challenged policy or practice is “arbitrary, artificial, and unnecessary to achieve a valid interest or legitimate objective such as a practical business, profit, policy consideration, or requirement of law.” HUD’s Implementation of the Fair Housing Act’s Disparate Impact Standard, 84 Fed. Reg. 42,854, 42,862 (Aug. 19, 2019) (to be codified at 24 C.F.R. pt. 100). Arguably, however, our arguments regarding the absence of legitimate reasons for targeting housing ads based on certain attributes might also satisfy this requirement.

112. HUD’s proposed rule for disparate impact claims permits defendants to rebut a prima facie case alleging that “the cause of a discriminatory effect is a model used by the defendant, such as a risk assessment algorithm” if the defendant

[p]rovides the material factors that make up the inputs used in the challenged model and shows that these factors do not rely in any material part on factors that are substitutes or close proxies for protected classes under the Fair Housing Act and that the model is predictive of credit risk or other similar valid objective.

Id. It is not at all obvious that attribute-based targeting involves the sort of predictive model covered by this provision or that allegations involving discriminatory attribute-based targeting amount to alleging that such a model is the “cause” of the discriminatory effect.

Consider first a platform's conduct in providing targeting attributes that are direct or apparent proxies for protected characteristics. Facebook has argued that, because its attribute-based tool is intended for general-purpose advertising, it cannot be liable merely for offering attributes that cannot be legally used to target housing ads.¹¹³ However, the platform's own conduct extends beyond designing an attribute-based tool that allows advertisers to select proxies for protected characteristics. It includes whether and how the platform marketed its attribute-based ad-targeting tool to housing advertisers and what, if anything, it does to inform housing advertisers or to prevent or discourage them from selecting prohibited targeting attributes. The totality of this conduct might be taken to "result in" discriminatory targeting even if merely including proxies for protected attributes in a general-purpose tool does not.

A platform's own conduct also includes defining, naming, and often inferring the attributes it offers. A platform might thus incur liability if it provides attributes that are defined, named, or inferred in a biased way that results in discriminatory targeting. For example, suppose an advertiser selects financial attributes that appear to be relevant to whether an ad recipient would be able to afford the housing on offer, but the inferences made by the platform's algorithms in assigning those attributes to users were systematically inaccurate for certain protected classes. While the advertiser's selection of those attributes would not appear discriminatory to an ordinary reader and would seem justified by legitimate business reasons, the platform might well be liable under any standard for constructing the attributes in a biased way.

If a platform successfully argues that its own conduct did not result in actionable housing discrimination, might it still be held liable for "failure to correct" its advertisers' attribute selections that resulted in discriminatory targeting? Unless courts adopt the expansive view of platform power and knowledge presented in Part III.B.1, it seems unlikely that platforms will be deemed to have the requisite power to control advertisers' attribute selections.

3. FHA Liability Arising from Lookalike Audience Selection

The lookalike audience tool shifts the burden of responsibility for audience selection almost entirely onto the platform's shoulders, often leaving it as the only potential discriminatory ad-targeting defendant. As a result, the

113. See Memorandum of Law in Support of Motion to Transfer Venue, or Alternatively to Dismiss Plaintiffs' First Amendment Complaint at 27, *Nat'l Fair Hous. All. v. Facebook, Inc.*, No. 1:18-cv-02689-JGK (S.D.N.Y. July 30, 2018), ECF No. 40 [hereinafter Facebook Brief] ("Facebook—as a website that publishes ads for many different types of goods and services, not just housing—is fundamentally different from a multiple-listing service or brokers' organization, which are run by brokers themselves for the sole and express purpose of offering and promoting housing opportunities.").

likelihood of platform liability grows, while the likelihood of advertiser liability diminishes.¹¹⁴

a. Advertiser Liability

Any advertiser liability for discriminatory targeting resulting from lookalike audience selection must arise from disparities in the base audience submitted by the advertiser. Because of the attenuated connection between the base audience and the demographics of the lookalike audience, however, such grounds for advertiser liability seem relatively weak. While the base audience provided by an advertiser helps determine the lookalike audience in a “but-for” sense, a court might—and perhaps should—conclude that it is the platform’s similarity algorithm, rather than the advertiser’s “own conduct,” that “results in” a disparately targeted lookalike audience.¹¹⁵

Facebook’s website currently says that a “Lookalike Audience is a way to reach new people who are likely to be interested in your business because they’re similar to your best existing customers.”¹¹⁶ In light of this promise, an ad-targeting plan that simply involves submitting a base audience of existing customers is unlikely to be perceived as problematic by an “ordinary reader.” Under a disparate impact analysis, targeting ads to “people who are likely to be interested in your business” also provides a seemingly legitimate reason for employing the tool. Facebook’s “similarity” metric is secret, leaving advertisers no grounds to anticipate how the base audiences they supply will influence the demographics of the lookalike audience. Indeed, based on Facebook’s statement, an advertiser might even hope that the lookalike audience tool would expand its target audience to include likely customers from currently underrepresented groups. In sum, advertisers neither control nor understand how a lookalike audience selection tool depends on the demographics of the base audience. It thus seems quite possible that courts would conclude that advertisers’ base audience submissions do not “result in” discriminatory ad targeting that emerges in the lookalike audience.

114. If courts adopt a disparate impact approach to automated ad targeting, HUD’s proposed changes to 24 C.F.R. § 100.500 seem most likely to affect claims involving lookalike audience targeting. Under the proposed rule, ad-targeting platforms might attempt to rebut a prima facie case by demonstrating that their similarity algorithms avoid using attributes that are “substitutes or close proxies for protected classes.” HUD’s Implementation of the Fair Housing Act’s Disparate Impact Standard, 84 Fed. Reg. at 42,862. If courts take the proposed rule’s standard seriously, however, rebuttal also requires defendants to demonstrate that their algorithm’s version of “similarity” metric is a “valid objective.” *Id.* That may be a trickier task if the algorithm produces disparately composed lookalike audiences.

115. 24 C.F.R. § 100.7 (2019).

116. *Create a Lookalike Audience*, FACEBOOK BUS., <https://www.facebook.com/business/help/465262276878947> [<https://perma.cc/8EDF-S6RM>] (last visited Oct. 6, 2019).

b. Platform Liability

Ad platforms may well be directly liable for discriminatory lookalike audience targeting based on their “own conduct” in designing and implementing algorithms that “result in” lookalike audience disparities. Such a platform might find it awkward to argue that the advertiser’s base audience selection is to blame because such an argument would suggest that the platform’s lookalike audience tool *tracks* the base audience’s balance of protected characteristics—an assertion platforms would presumably want to avoid. Moreover, a general-purpose tool argument seems unpersuasive when applied to a lookalike audience tool that promises to use the platform’s user data and algorithm to find promising customers for an advertiser’s business.

In the unlikely event that a court decides that an advertiser’s submission of a disparate base audience resulted in discriminatory targeting while the ad-targeting platform’s algorithm design did not, plaintiffs would have a particularly strong argument for holding the platform liable for “failure to correct” the discriminatory targeting. Whereas the general “failure to correct” argument described in Part III.B.1 would define what an ad-targeting platform “should have known” by *the data available to it*, a lookalike audience algorithm must *actually analyze* the attributes of both base and lookalike audience members in order to assess similarity. Moreover, Facebook’s promise to create lookalike audiences that are “likely to be interested in your business”¹¹⁷ suggests that its lookalike audience algorithm analyzes the content of the ad and thus that it “should know” when it is dealing with a housing ad.

4. Possible Impact of the Consent Decree

The changes required by the March 2019 consent decree will lower the overall chances of an underlying FHA violation in several ways. Eliminating the options to target housing ads by zip code, gender, age, and other attributes highly correlated with protected classes makes it less likely that advertisers will use Facebook’s tools to discriminate. Likewise, limiting the factors the lookalike audience algorithm considers in assessing similarity may result in less disparate results. It is by no means clear, however, that these changes will substantially reduce discriminatory housing ad targeting or protect Facebook from future liability.¹¹⁸

Many unprotected targeting attributes will continue to be available to housing advertisers even though they appear to be irrelevant to reasonable housing ad targeting. If selecting such attributes produces disparate targeting, advertisers might still have trouble providing legitimate reasons for using them and Facebook may not be any more persuasive in articulating legitimate reasons for including them as options for housing advertising. It is similarly hard to imagine legitimate reasons, in the housing context, for

117. *Id.*

118. The changes may be more protective, however, if courts decide to apply a disparate impact analysis and HUD’s proposed changes to 24 C.F.R. § 100.500 are adopted.

relying on many of the attributes that the settlement permits Facebook to continue using to create lookalike audiences.

IV. CDA 230 IMMUNITY AND AUTOMATED AD TARGETING

As discussed in Part I.B, aside from making mostly unsuccessful arguments for a narrower interpretation of “publisher,” plaintiffs have focused on two types of arguments against CDA 230 immunity: (1) that the claims at issue do not treat the defendant as a speaker or publisher; and (2) that the service provider is an information content provider responsible for the development of actionable content by materially contributing to it (often through the design of its services) or inducing it. This Part considers how these arguments are likely to fare against CDA 230 assertions by ad-targeting platforms.

A. Ad Targeting as a Publisher Activity?

In general, targeting advertising is quite likely to be considered a “publisher” activity under CDA 230. As Facebook argues in litigation:

Publishers routinely allow advertisers to target ads and assist them in doing so. For instance, newspapers sell advertising space to sporting goods store[s] in the sports section and broadcast networks sell airtime to advertisers on different channels during different shows depending on the preferred audience. The publication of ads entails more than displaying them; it also entails allowing advertisers the ability to reach their target audience.¹¹⁹

The NFHA counters that while “[a] traditional ‘publisher’ is a website that ‘reviews material submitted for publication, perhaps edits it for style or technical fluency, and then decides whether to publish it,’”¹²⁰ “Facebook’s conduct goes far beyond this ‘reviewing’ and ‘editing’”¹²¹ and its ad-targeting service thus is not a “publisher” function under the CDA.¹²² Facebook correctly notes, however, that “[c]ourts have consistently held in analogous contexts that ad targeting is a traditional publisher function protected by the CDA.”¹²³

While it seems likely that courts would credit Facebook’s argument that ad targeting is generally a “publisher” activity, plaintiffs might nonetheless argue that CDA 230 does not preclude liability based on a platform’s “failure to correct” advertiser discrimination because such claims do not treat the

119. Facebook Brief, *supra* note 113, at 24–25.

120. Memorandum of Law in Opposition to Defendant’s Motion to Transfer Venue or Alternatively to Dismiss Plaintiffs’ First Amended Complaint at 15, Nat’l Fair Hous. All. v. Facebook, Inc., No. 1:18-cv-02689-JGK (S.D.N.Y. Aug. 24, 2019), ECF No. 49 [hereinafter NFHA Reply Brief] (citing *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1102 (9th Cir. 2009)).

121. *Id.*

122. *Id.*

123. Facebook Brief, *supra* note 113, at 25 (first citing *Jane Doe No. 1 v. Backpage.com LLC*, 817 F.3d 12, 16 (1st Cir. 2016); then citing *Fair Hous. Council v. Roommates.com, LLC*, 521 F.3d 1157, 1169 (9th Cir. 2008) (en banc)).

platform as a publisher. In *Espinoza v. County of Orange*, a court recently denied an employer CDA 230 immunity for claims alleging an employer's "failure to investigate and resolve" discriminatory workplace harassment,¹²⁴ where the harassment involved derogatory blog postings made by coworkers using work computers.¹²⁵ The court noted that the "plaintiff is not arguing defendant is the publisher of the blog postings" and that the "defendant's breach was not based on its employees' use of their work computers but on its own failure to investigate and resolve the problem."¹²⁶ An FHA plaintiff could advance a similar argument that liability for "failure to correct" discriminatory ad targeting does not treat the ad platform as a publisher.

Under current CDA 230 precedent, courts are unlikely to be convinced of the analogy. *Espinoza* alleged that his employer should have resolved the harassment by policing its employees' use of work computers more effectively, an action unrelated to publishing.¹²⁷ Correcting discriminatory ad targeting, however, necessarily means either directly modifying the target audience or changing how it is selected, actions that seem quite similar to the steps that a conventional publisher would take to avoid publishing discriminatory ads.

Courts are thus likely to conclude that ad targeting is at least generally a "publisher" activity and to reject plaintiffs' arguments that "failure to correct" claims in particular are beyond the scope of CDA 230. For the most part, then, plaintiffs' success will depend on whether ad-targeting platforms are treated as "information content providers" under CDA 230.

B. Ad-Targeting Platforms as Information Content Providers?

A party engaged in nominally publisher-type activities is not immune under CDA 230 to the extent it acts as an "information content provider" "responsible, in whole or in part, for the creation or development of information provided through [an] interactive computer service."¹²⁸ In the NFHA litigation, Facebook argued that it is not an "information content provider" of discriminatorily targeted advertising because "[p]laintiffs do not, and cannot, allege that Facebook 'materially contribut[ed] to [the] alleged unlawfulness' of the content or 'assisted in the development of what made the content unlawful.'"¹²⁹ The NFHA and HUD responded that Facebook is an "information content provider" because: (1) it "seizes upon

124. *Espinoza v. County of Orange*, No. G043067, 2012 WL 420149, at *9 (Cal. Ct. App. Feb. 9, 2012); *see also* *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 851 (9th Cir. 2016) (finding no immunity where an "obligation to warn could have been satisfied without changes to the content posted by the website's users"). *But see* *Nat'l Ass'n of the Deaf v. Harvard Univ.*, 377 F. Supp. 3d 49, 66 (D. Mass. 2019) (CDA 230 applies even "when a plaintiff claims disability discrimination based on a lack of access rather than on the content of speech.").

125. *Espinoza*, 2012 WL 420149, at *2–3.

126. *Id.* at *9.

127. *See id.* at *8–9.

128. 47 U.S.C. § 230(f)(3) (2012); *see supra* Part I.B.

129. Facebook Brief, *supra* note 113, at 19 (quoting *FTC v. LeadClick Media, LLC*, 838 F.3d 158, 174 (2d Cir. 2016)).

benign content posted by its users—not merely sex and location but also personal interests, demographic information, and browsing history—to actively classify its users into categories” that enable discriminatory ad targeting; (2) its “ad utilities . . . invite housing providers to express unlawful demographic and other audience preferences,” making “Facebook and the advertiser . . . co-developers of each targeted ad”;¹³⁰ and (3) it develops the “information” contained in the attributes offered by its targeting platform.¹³¹

These arguments illustrate the difficulty in applying CDA 230 to Facebook’s ad-targeting platform and similar smart services. Facebook is clearly correct that its targeting tools have nothing to do with developing the third-party ad content. While HUD and the NFHA make plausible arguments that Facebook is, nonetheless, materially responsible for discriminatory ad targeting, they have some difficulty shoehorning this responsibility into CDA 230’s “information content provider” definition, which naturally focuses on contributions to “content.” These tensions, also reflected in judicial opinions grappling with the “information content provider” definition, reflect deep issues with the structure and interpretation of CDA 230, which we analyze further below.

For now, we analyze how FHA claims against ad-targeting platforms would fare under the CDA 230 test preferred by the plaintiffs in the NFHA litigation, which evaluates whether a platform is “responsible, in whole or in part, for the creation or development” of discriminatory ad targeting, using the approach developed in *Roommates.com*.¹³² Moreover, whatever the eventual scope of “failure to correct” liability under the FHA, only an ad-targeting platform’s “own conduct” can turn it into an “information content provider” that is ineligible for CDA 230 immunity.

1. Attribute-Based Audience Selection

The *Roommates.com* analysis contrasted “neutral tools” with online services that made “material contributions” to third-party actionable conduct. In litigation over its ad-targeting tools, Facebook emphasizes that “[a] material contribution to the alleged illegality of content [under the CDA] . . . means being responsible *for what makes the displayed content allegedly unlawful*”¹³³ and contends that it does nothing to require or encourage advertisers to select attributes that result in disparate audiences.¹³⁴ It argues further that its “provision of neutral tools that . . . *may* be used by some *advertisers* for an unlawful purpose falls squarely within the scope of CDA immunity.”¹³⁵

130. HUD Brief, *supra* note 6, at 18, 21, 22.

131. See NFHA Reply Brief, *supra* note 120, at 16.

132. *Id.* at 9.

133. Defendant’s Notice of Motion and Motion to Dismiss First Amended Complaint; Memorandum of Points and Authorities in Support Thereof at 9, *Mobley v. Facebook, Inc.*, No. 5:16-cv-06440-EJD (N.D. Cal. Apr. 3, 2017), ECF No. 34 (quoting *Jones v. Dirty World Entm’t Recordings LLC*, 755 F.3d 398, 410 (6th Cir. 2014)).

134. *Id.* at 19.

135. Facebook Brief, *supra* note 113, at 2.

Plaintiffs, in contrast, argue that Facebook's active involvement in defining, developing, and assigning attributes to users and offering them to advertisers materially contributes to discriminatory targeting,¹³⁶ particularly when the available attributes include protected characteristics.¹³⁷ Analogizing to *Roommates.com*, they argue further that "[w]hen a business enterprise extracts such information from potential customers [about their protected statuses] as a condition of accepting them as clients, . . . the enterprise is responsible, at least in part, for developing that information."¹³⁸

These opposing contentions demonstrate that "neutral tools" and "material contributions" are not simply two sides of the same coin. Attribute-based targeting is *neutral* in that it *can* be used to target housing (and other) advertising in either a discriminatory or nondiscriminatory fashion. Nonetheless, Facebook's design and definition of attributes that proxy for protected characteristics do contribute materially to the development of any discriminatory ad targeting that makes use of them.

It is hard to predict how a court would come out on this issue. Given the general tendency of courts to favor immunity, Facebook may have the better of the argument¹³⁹ and ad-targeting platforms may ordinarily be shielded from liability for discriminatory attribute-based targeting.

2. Lookalike Audience

A platform's role in designing the algorithm used to create a lookalike audience tool undoubtedly contributes materially to whether disparate targeting results. Whether the lookalike audience tool is nonetheless a "neutral tool[]" protected by the CDA,¹⁴⁰ as Facebook contends, depends on what "neutral" means in this context. If "neutral" describes a tool that was not designed to *promote* discrimination, lookalike audience targeting may be neutral. If, on the other hand, "neutral" means that the user, not the tool designer, *controls* whether the outcome is discriminatory or nondiscriminatory, lookalike audiences targeting is definitely not neutral. As discussed in Part III.B.2.b, while they contribute a "base audience," advertisers have essentially no say whatsoever in whether the lookalike audiences used to target their ads are disparately targeted. While it remains unclear how courts will resolve the "neutral tool" question, plaintiffs' arguments against CDA 230 immunity for ad-targeting platforms are certainly strongest for the lookalike audience tool.

136. NFHA Reply Brief, *supra* note 120, at 16.

137. *See, e.g., id.* at 3.

138. *Id.* at 10–11.

139. *See supra* Part I.B.

140. Facebook Brief, *supra* note 113, at 21 (quoting *Fair Hous. Council v. Roommates.com, LLC*, 521 F.3d 1157, 1171 (9th Cir. 2008) (en banc)).

C. *Putting It All Together: Can Plaintiffs Recover Against Ad-Targeting Platforms?*

Putting Parts III and IV together, overall, we would roughly expect that: (1) failure to correct claims are likely to fail because of CDA 230 immunity; (2) attribute-based targeting claims are somewhat likely to be viable under the FHA but reasonably likely to fail nonetheless because of CDA 230; and (3) lookalike audience-based claims are the most likely to succeed under both the FHA and CDA 230. If ad-targeting platforms are successful in claiming CDA 230 protection, plaintiffs may have no recourse because advertisers are unlikely to be held liable.

V. SMART SERVICES, SECONDARY LIABILITY, AND CDA 230

Though not ordinarily framed in these terms, the tensions and trade-offs underlying the controversy about CDA 230 immunity are helpfully understood by reference to debates about various forms of secondary liability.¹⁴¹

A. *Secondary Liability Paradigms*

Secondary liability often aims to enlist the assistance of large, institutionalized, deep-pocketed players in enforcing laws when enforcement against those who are directly liable is ineffective or costly.¹⁴² It can also be intended to rein in the behavior of intermediaries who facilitate the underlying illegality or amplify its harms. Secondary liability routinely provokes controversy about who is an appropriate defendant and what level of culpability or mental state should be required. When the boundaries between legal and illegal underlying behavior are not bright,¹⁴³ there is often fear that secondary liability defendants, who may have more to lose and more litigation risk than direct liability defendants, will overcompensate. The extent to which overcompensation is either likely or problematic depends on the context and the type of secondary liability involved.¹⁴⁴

Secondary liability can be divided into four basic types: (1) what we will term “monitor and control” liability, exemplified by vicarious liability for those with a right, ability, or duty to control the actions of those directly liable; (2) notice-based liability; (3) liability for inducement of a third party’s actionable conduct; and (4) contributory liability based on facilitating or contributing to actionable conduct. Secondary liability often requires some

141. Note that the immunity framework provided by CDA 230 has been referenced in the context of the Digital Millennium Copyright Act (DMCA) and the issues with a safe harbor approach to internet service provider (ISP) copyright immunity given the state of the modern internet. See, e.g., Katherine Burkhart, Note, *Mavrix v. LiveJournal: Unsafe Harbors in the Age of Social Media*, 33 BERKELEY TECH. L.J. 1015, 1029–33 (2018).

142. See Wu, *supra* note 16, at 302–04.

143. This is obviously true of speech-related claims but also for claims involving intellectual property infringement where there are gray lines between socially beneficial uses and infringement.

144. See generally Wu, *supra* note 16; Grimmelmann, *supra* note 16.

level of knowledge or intent; the appropriate level for a given context is debatable. Courts and commentators have grappled extensively with secondary liability for online copyright infringement, which is a useful comparator here since some (though not all) of the underlying policy issues are similar.¹⁴⁵

1. Publishers, Distributors, and Secondary Liability

Because defamation doctrine was the mental model underlying CDA 230, we use its relationship to secondary liability as a framework for introducing secondary liability. Defamation involves the communication of content that is false and harmful to reputation.¹⁴⁶ The doctrine distinguishes between original speakers, (re-)publishers, and those who “merely [make] available to another equipment or facilities that he may use himself for general communication purposes.”¹⁴⁷

Original speakers and publishers are ordinarily held liable even if they are only negligent,¹⁴⁸ meaning they did not act “reasonably in checking on the truth or falsity or defamatory character of the communication before publishing it.”¹⁴⁹ The duty to check varies contextually. For example, professional publishers, such as newspapers, are judged by the “skill and experience normally possessed by members of that profession.”¹⁵⁰ While publisher liability is framed as direct liability, it is effectively a form of secondary liability premised on actionable third-party content.

Publisher liability is one form of “monitor and control” liability. Like vicarious liability, it subjects certain secondary parties to affirmative duties to monitor others’ behavior, but the negligence standard makes the duty less absolute. Olivier Sylvain’s suggestion to require online service providers to make “good faith” efforts to screen for discrimination is another sort of “monitor and control” standard.¹⁵¹ Depending on the stringency of the standard and the cost and feasibility of distinguishing actionable from permissible third-party behavior,¹⁵² “monitor and control” liability can incentivize overenforcement, including “collateral censorship.”

Distributors, such as booksellers and news vendors, are liable only if they know or have reason to know specific content is defamatory.¹⁵³ Distributor liability is a form of notice-based liability. The Digital Millennium

145. See Grimmelmann, *supra* note 16.

146. RESTATEMENT (SECOND) OF TORTS §§ 558–59 (AM. LAW INST. 1976).

147. *Id.* § 581 cmt. b. This terminology can be a bit confusing because the communication element of defamation is also called “publication.” *Id.* § 558. References to “publisher liability,” as well as interpretations of CDA 230’s use of the term, however, connote republishers of third-party content.

148. For exceptions, see RESTATEMENT (SECOND) OF TORTS §§ 583–612 (AM. LAW INST. 1976). Most notably, for statements about public figures, see *id.* § 580A.

149. *Id.* § 580B cmt. g.

150. *Id.*

151. Sylvain, *supra* note 25.

152. Wu, *supra* note 16; Grimmelmann, *supra* note 16.

153. RESTATEMENT (SECOND) OF TORTS § 581 (AM. LAW INST. 1976); *see id.* cmt. b.

Copyright Act's (DMCA) "safe harbor" for online service providers who adopt a statutory "notice and takedown" regime provides another example.¹⁵⁴ While notice-based liability relieves secondary parties of monitoring duties, it can still incentivize overenforcement, especially when it is difficult or costly to evaluate whether a user's conduct is actionable.¹⁵⁵ Service providers may then respond by censoring conduct too readily. Especially in the online context, this may incentivize dubious notices, which ratchet up service provider costs and create a feedback loop. As James Grimmelman puts it, while defamation "is a doctrinal swamp where cases often turn on subtle nuances of meaning," copyright seems like a promising context for online notice-based liability, since "the prima facie question of whether a particular piece of content is or is not a nearly identical copy of a particular copyrighted work is something a platform can delegate to a hashing algorithm."¹⁵⁶ Nonetheless, the DMCA's notice-and-takedown regime is widely criticized for allowing "take-down" without adequate proof of the underlying infringement.¹⁵⁷

The degree to which notice-based liability fosters overenforcement also depends on the duties imposed on secondary parties. A service provider is only eligible for the DMCA safe harbor, for example, if it "responds expeditiously to remove" allegedly infringing material "upon notification."¹⁵⁸ Although users can dispute the takedown later, based on fair use and other defenses, the DMCA essentially codifies overenforcement. FHA "failure to correct" liability is also a version of notice-based liability, given its "knew or should have known" trigger. Neither the standard nor the context invites the knee-jerk overenforcement mandated by the DMCA, however. Because of the variations possible among both notice-based and "monitor and control" regimes, and particularly because of the issue of dubious notices in the online context, there is no single answer to which sort of regime creates greater overenforcement incentives.

2. Contributory Liability

Contributory liability¹⁵⁹ attaches to a secondary party who contributes to or facilitates another party's actionable behavior but whose conduct is not directly actionable. To avoid ensnaring innocent behavior, contributory liability regimes often exempt providers of general-purpose components, equipment, or facilities and impose mental-state standards beyond mere

154. 17 U.S.C. § 512 (2012).

155. Amanda Reid, *Considering Fair Use: DMCA's Take Down & Repeat Infringers Policies*, 24 COMM. L. & POL'Y 101, 105 (2019) ("In this haze, ISPs are incentivized to over-block and err on the side of removing content—including lawful content.").

156. Grimmelman, *supra* note 16 (making a similar point about child pornography as well).

157. *See, e.g.*, Reid, *supra* note 155, at 123–24.

158. 17 U.S.C. § 512(c)(1)(C).

159. The term is sometimes used to refer to both what we call "contributory liability" and what we call "inducement." We take our terminology from the patent statute, which distinguishes them clearly. 35 U.S.C. § 271(b)–(c) (2012).

negligence. Although defamation doctrine does not incorporate a fully realized version of contributory liability, it similarly provides immunity for those who merely make “equipment or facilities” available to others for “general communication purposes.”¹⁶⁰

Contributory liability standards vary. Contributory copyright infringement, for example, requires proof that the defendant provides a technology or service that materially contributes to another’s infringement and is not “capable of substantial noninfringing uses.”¹⁶¹ Patent law requires a showing that the defendant’s technology was “especially made or especially adapted for [infringing] use.”¹⁶²

Contributory liability is distinguished from both “monitor and control” and noticed-based liability in that it does not require secondary parties to do anything about others’ actionable conduct except to refrain from contributing to it. The doctrine thus attempts to avoid overenforcement issues. Depending on the specific standards for actionable contributions and mental-state requirements that are adopted,¹⁶³ different contributory liability regimes may create somewhat different balances between overenforcement and harm reduction, which reflect different policy trade-offs.

The exemption for general-purpose services has a more universal justification because imposing liability on such a service in one substantive arena may create negative externalities in others. For this reason, it is reasonable to include a general-purpose service exemption in any contributory liability regime.

3. Inducement Liability

Inducement liability, which is incurred by encouraging others to engage in actionable conduct, has no analog under traditional defamation doctrine. It “premises liability on purposeful, culpable expression and conduct” and neither “mere knowledge of infringing potential or of actual infringing uses” nor incidental acts “such as offering customers technical support or product updates” are sufficient.¹⁶⁴ Inducement liability is broader than contributory liability in some respects because it does not require a material contribution to others’ conduct by the platform.¹⁶⁵

160. RESTATEMENT (SECOND) OF TORTS § 581 cmt. b (AM. LAW INST. 1976).

161. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984).

162. 35 U.S.C. § 271(c) (2012).

163. In intellectual property law, for example, there have been longstanding debates about the contours of the mental state requirements for contributory and inducement liability and several Supreme Court decisions address the issue. *See Commil USA, LLC v. Cisco Sys., Inc.*, 135 S. Ct. 1920 (2015) (patent infringement); *Global-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754 (2011) (patent infringement); *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005) (copyright infringement). *See generally* Charles W. Adams, *Indirect Infringement from a Tort Law Perspective*, 42 U. RICH. L. REV. 635 (2008) (situating intellectual property’s secondary liability doctrines within tort law).

164. *Grokster*, 545 U.S. at 937.

165. *See id.*

Inducement liability's focus on the secondary defendant's own actions and mental state arguably avoids creating overenforcement incentives. The U.S. Supreme Court opined that copyright's inducement liability "does nothing to compromise legitimate commerce or discourage innovation having a lawful promise."¹⁶⁶ That view is probably too rosy. The fuzziness of inducement liability's boundaries may create some incentive to avoid gray areas. The line between inducing actionable and nonactionable behavior (say, defamation or profane criticism) is not always bright. Proving mental state is notoriously difficult.¹⁶⁷ Moreover, factfinders might assess the defendant's own statements and the design of its service through the prism of its users' conduct. The extent to which inducement liability will produce risk-averse behavior by service providers depends on the context, as does the extent to which society should care.

B. CDA 230 Through a Secondary Liability Lens

CDA 230's defamation-focused design contained within it the seeds of problems that have borne unfortunate fruit, especially as smart services have emerged. Those seeds were of two varieties: first, CDA 230's drafters did not anticipate the wide variety of conduct and legal claims it could potentially immunize; second, CDA 230 adopted defamation doctrine's all-or-nothing approach to secondary liability, ignoring the possibility that the socially preferable secondary enforcement regime might depend on the substantive context.

1. CDA 230 and Service Provider Conduct

CDA 230 envisions only three roles for online service providers: publisher¹⁶⁸ of third-party information content, information content provider, and, implicitly, general-purpose facility. Relatedly, CDA 230 reflects an assumption that actionable *content* is the only source of liability for publisher activities.

Online service providers now engage in a broad variety of activities that CDA 230 was not designed to handle. Automated filtering, sorting, and prediction have vastly changed the scope of online publishers' activities, which can now create or contribute to liability that is not based on actionable content. CDA 230 holds publishers immune unless they create or develop actionable content.¹⁶⁹ As a result, the statute now often immunizes publishers from liability for their own directly actionable conduct.

166. *Id.*

167. See generally Samuel W. Buell, *Good Faith and Law Evasion*, 58 UCLA L. REV. 611 (2011) (discussing some of the complexities involved in proving state of mind).

168. CDA 230's reference to a "publisher or speaker" is redundant, since a "speaker" of third-party content is also generally a "publisher." The inclusion of "speaker" seems to play no independent role.

169. See, e.g., *Huon v. Denton*, 841 F.3d 733, 741–42 (7th Cir. 2016); *FTC v. LeadClick Media, LLC*, 838 F.3d 158, 173 (2d Cir. 2016).

Our ad-targeting case study illustrates this problem. Facebook argues that it is immune from liability for creating and offering targeting attributes and for its lookalike audience algorithm essentially because they are publisher activities that do not contribute to content creation.¹⁷⁰ Courts may or may not reject that argument in favor of a *Roommates.com* approach. Either outcome begs the question of why an online service provider's direct liability for discriminatory ad targeting should turn on whether the claim treats the provider as the publisher of another party's information content.

CDA 230 likely immunizes ad-targeting platforms from the FHA's effectively secondary "failure to correct" liability. Whether this particular form of notice-based liability would impose excessive costs on a smart ad-targeting service is a debatable matter of housing discrimination policy. Whether the targeting platform fits CDA 230's information content provider definition seems completely irrelevant to that policy question.

The problem illustrated here is that, even though it seems sensible to characterize ad targeting as a publisher activity, allegations of discriminatory ad targeting have nothing to do with content development. CDA 230 was simply not designed or intended to handle situations in which a service provider's *activities* as a publisher are actionable but the published *content* is not.

2. CDA 230's All-or-Nothing Approach to Secondary Liability

Like defamation doctrine, CDA 230 sets up an all-or-nothing choice between "monitor and control" liability and complete immunity. Courts have struggled to find more nuanced ways to apply CDA 230 to today's increasingly smart online service providers, but it is a difficult struggle because the definition of "information content provider" is essentially the only available statutory hook.

a. Shoehorning in Contributory Liability

Roommates.com essentially shoehorned a contributory liability approach into the information content provider definition.¹⁷¹ The case interpreted "responsibility" for developing online information in terms of whether the website's design "materially contributed" to its users' discriminatory advertising or was a "neutral tool."¹⁷² This test is, of course, familiar from contributory liability.

While contributory liability may often be a good approach to online service provider liability, CDA 230's "information content provider" definition is an awkward place to insert it. The awkwardness derives not only from the fact that the definition focuses on content but also from its statutory role in defining when defendants are directly liable. Contributory liability standards

170. Facebook Brief, *supra* note 113, at 24–25.

171. Fair Hous. Council v. *Roommates.com, LLC*, 521 F.3d 1157, 1172 (9th Cir. 2008) (en banc).

172. *Id.*

tend to be different from the underlying direct liability standards precisely because the defendant's liability is derivative. A contributory liability analysis is thus a muddled way to decide whether to immunize a defendant from direct liability.

Moreover, the externality-based rationale for immunizing general-purpose services ("neutral tools") from contributory liability does not apply to direct liability stemming from actionable *use* of such tools. For direct liability based on an online service's *design*, however, it makes sense to provide immunity for general-purpose services that could not reasonably be redesigned to avoid direct liability without sacrificing substantial legitimate uses.

The ad-targeting case study is again illustrative. The FHA imposes direct liability when a defendant's own conduct results in a discriminatory housing practice, as evaluated under the applicable standards. Those standards incorporate their own, quite rigorous, tests for distinguishing legitimate activities from actionable discrimination. Those tests may not align with a "neutral tool" exemption. For example, proffering a "legitimate reason" for disparate impact requires more than mere neutrality. Facebook argues that its attribute-based targeting tool is "neutral" because it offers the same attributes to all potential advertisers.¹⁷³ Such neutrality need not imply that an ad-targeting service has a legitimate reason for offering protected attributes to housing advertisers. It may simply mean that the service's designers did not bother to account for antidiscrimination laws. Immunizing an ad-targeting service from FHA liability simply because it is "neutral" seems perverse, particularly if the design could have accounted for antidiscrimination laws while maintaining the service's viability for other sorts of advertisers. The earlier discussion, along with the consent decree, suggests that this was the case for Facebook.

b. Putting Inducement Back on the Table

Courts have considered inducement-like arguments against CDA 230 immunity but mostly have not been persuaded by them.¹⁷⁴ This is probably the right result under the current statute, despite the fact that the wrongs alleged in some cases clearly arise from the service provider's inducement of actionable conduct by its users. Inducement liability is not appropriate in all arenas because it can induce some overcompliance, but that is no reason to insist on CDA 230's all-or-nothing Hobson's choice. How worried do we really need to be about overdetering service providers from bellying right up to the line of actively inducing housing discrimination or fraud or child sex trafficking? Within the strictures of the First Amendment, those trade-offs are matters for political debate. By enacting FOSTA, which creates inducement liability, Congress answered that question for prostitution and

173. Facebook Brief, *supra* note 113, at 27, 30.

174. See *supra* notes 40–41 and accompanying text.

sex trafficking. FOSTA is controversial,¹⁷⁵ but that makes our point. The adoption of inducement liability in particular substantive arenas should be on the table for debate.

c. Institutional Competence

Many substantive liability regimes incorporate secondary liability standards (whether or not so-named) that are much more nuanced than CDA 230's all-or-nothing approach. Regardless of whether one agrees with the results, those regimes presumably reflect balances worked out in light of policy concerns salient in those arenas. CDA 230 runs roughshod over those balances whenever online service providers are involved. It is still true that balances struck for offline secondary liability may be upset online and also that externality concerns may be more urgent simply because general-purpose services may be more common. Nonetheless, today's online services are sufficiently various that it is no longer plausible, particularly given the rise of smart services, that CDA 230's all-or-nothing balance is appropriate for all of them.

CONCLUSION

To conclude, we offer a proposal for reform of CDA 230 based on the above analysis.

1) *Definition of "treated as a publisher."* CDA 230 should be amended to clarify that a party is not "treated as the publisher or speaker of any information provided by another information content provider" unless liability is premised primarily on the actionable nature of that third-party content. This change preserves the sort of immunity from publisher liability that the drafters of CDA 230 had in mind.¹⁷⁶ We also recommend that the phrase "publisher or speaker" be modified to "publisher or distributor," adopting current judicial interpretation.

2) *Limited immunity for services capable of substantial nonactionable uses.* The previous amendment would clarify that online service providers are generally liable for direct and secondary claims not primarily premised on actionable third-party content. It is generally appropriate to defer to the substantive policy trade-offs made by legislators and relevant agencies. We see no justification, particularly in an era of smart services, for a blanket assumption that conforming to substantive legal constraints will always be unduly burdensome for online service providers.

We recognize, however, that some substantive laws and regulations on the books will not adequately account for the potential external social costs of holding providers of general-purpose online services liable. We therefore

175. See, e.g., Lura Chamberlain, Note, *FOSTA: A Hostile Law with a Human Cost*, 87 FORDHAM L. REV. 2171, 2188–89 (2019) ("Opponents contend that FOSTA could reach any UISP shown to have hosted actionable material related to sex trafficking even if the UISP had neither knowledge of the content nor the intent to assist sex traffickers.").

176. See *supra* Part I.A.

propose adding a provision to CDA 230 conferring immunity on providers of online services capable of substantial nonactionable uses when two conditions are satisfied: (1) liability is based on the design of the service; and (2) the service cannot reasonably be designed to avoid liability while retaining substantial nonactionable uses. The second condition reflects the fact that software-based services can often be relatively cheaply and easily customized for different applications. It is also good policy to incentivize service providers to incorporate such reasonable measures into their designs. We would apply the proposed limited immunity to both direct and secondary liability.

The *Roommates.com* and ad-targeting cases can be used to illustrate how this proposal would work in practice. *Roommates.com*'s design may or may not have had substantial nonactionable uses but, in any event, could easily have been modified to avoid liability under the FHA simply by removing protected characteristics.¹⁷⁷ The designers were at best reckless with regard to the site's potential to promote housing discrimination and there seems to be no normative justification for immunizing them from liability for their design. While the Ninth Circuit arrived at the right result, our proposal would provide a much cleaner path to that outcome.

Facebook's ad platform is certainly capable of substantial nonactionable uses, but reasonable design changes along the lines of those in the consent decree would likely have enabled Facebook to reduce its potential liability for discriminatory ad targeting without impacting the platform's nonactionable uses.

3) *Immunity from secondary liability not based on publishing actionable third-party content.*¹⁷⁸ We do not propose complete immunity from all substantive secondary liability provisions. Especially in the era of smart services, online service providers may have the capacity to cope with many secondary liability regimes, even if they take "monitor and control" or notice-based approaches. Deference to the policy trade-offs enshrined in substantive secondary liability regimes is generally appropriate, but many laws on the books may not have considered how the online environment might affect those policies. We therefore propose modifying CDA 230 as follows:

- a) Secondary liability provisions based on a defendant's contribution to, facilitation of, or failure to monitor actionable user behavior would be preempted and replaced by a contributory liability regime that combines a "material contribution" requirement with a "knew or should have known" mental state, unless and until regulators redesigned or reaffirmed them for online services. This default contributory liability regime would, for example, replace substantive "monitor and control" regimes such as the FHA's "failure to correct" provisions that had not been redesigned or reaffirmed.

177. See *Roommates.com*, 521 F.3d at 1167–68.

178. This provision applies even if it is framed as "direct liability," such as the liability for "failure to correct" under 24 C.F.R. § 100.7(a)(ii)–(iii).

b) Online service providers would be immune from substantive inducement liability regimes, unless and until regulators redesigned or reaffirmed their applicability. While our proposal puts inducement liability back on the table, we do not think inducement liability is an appropriate default, given the First Amendment and collateral censorship concerns.

c) These rules would apply to both federal and state statutes and regulations.

As an example, an ad-targeting platform would be immune from liability under the FHA's current "failure to correct" provisions because they are not specifically applicable to online service providers. The service provider would nonetheless be liable under the default contributory liability regime if it knew or should have known that its ad-targeting platform materially contributes to discriminatory ad targeting. In the future, HUD could use notice-and-comment rulemaking to specifically extend the "failure to correct" regime to online service providers. Moreover, our proposed version of CDA 230 would not immunize online service providers from liability if Congress were to, if it so desired, add a provision to the FHA providing liability for actively inducing discriminatory uses of automated ad-targeting tools.