## COMMENT

# ENIGMA MACHINES: DEEP LEARNING ALGORITHMS AS INFORMATION CONTENT PROVIDERS UNDER SECTION 230 OF THE COMMUNICATIONS DECENCY ACT

VINCENT DUMAS

Since Congress enacted the Communications Decency Act in 1996, the technical sophistication of algorithms and data collection on the internet has developed exponentially. Congress intended Section 230 of the CDA to provide protection to internet companies for censoring obscene and offensive content on their websites. However, even as the state of web technology has snowballed, courts have increasingly double downed on their initial, expansive construction of Section 230 immunity for algorithms. Most recently, several circuits found blanket Section 230 immunity for internet companies' algorithmic recommendations of ISIS recruitment videos to users prone to radicalization.

This Comment argues that courts' original construction of Section 230 in the context of surfacing algorithms is divorced from the statutory text and maladapted to modern realities. Courts' use of "neutral tool" analysis in deciding Section 230 immunity has proven unworkable in the face of increasingly sophisticated machine learning algorithms and mass data collection. In lieu of neutral tool analysis, this Comment proposes an alternative framework—the discretionary test—for evaluating whether information content has been created or developed under Section 230. The narrower scope of Section 230 immunity under the discretionary test would better allow litigants the opportunity for discovery and fair recovery against internet service providers employing machine learning algorithms. Finally, this Comment discusses the other defenses to liability that internet service providers employing deep learning algorithms would still enjoy should Section 230 algorithmic immunity be limited to its proper scope.

## INTRODUCTION

Between 1996, when Congress enacted the Communications Decency Act (CDA),[1] and the present day, societal use of the internet erupted.[2] During that time, the amount of data sent over the internet globally mushroomed by over 20,833,200 percent.[3] That increase in internet usage was driven by the astounding pace of technological advancement,[4] with key computer components doubling in capacity every year.[5] This seismic shift in technology that occurred after Congress enacted the CDA now appears as but a dewdrop before the downpour that is to develop in the coming decades.[6] On the web, simple, deterministically populated HTML pages have been replaced by dynamic sites populated with content from complex machine learning algorithms

1. Telecommunications Act of 1996, Pub. L. No. 104-104, tit. 5, 110 Stat. 133-43 (codified as amended at 47 U.S.C. § 230).

2. *Compare Americans Going Online…Explosive Growth, Uncertain Destinations*, PEW RSCH. CTR. (Oct. 16, 1995), https://www.pewresearch.org/politics/1995/10/16/americans-going-online-explosive-growth-uncertain-destinations [https://perma.cc/CW4J-H945], *with Internet/Broadband Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), https://www.pewresearch.org/internet/fact-sheet/internet-broadband [https://perma.cc/K8JB-4PTK]. *See generally Internet Growth Statistics*, INTERNET WORLD STATS, https://www.internetworldstats.com/emarketing.htm [https://perma.cc/W922-4DLS] (last visited Nov. 8, 2022).

3. *See Crossing Borders*, WORLD BANK, https://wdr2021.worldbank.org/stories/crossing-borders [https://perma.cc/CXB8-TRWZ] (last visited Oct. 30, 2022); Arielle Sumits, *The History and Future of Internet Traffic*, CISCO: SP360: SERVICE PROVIDER (Aug. 28, 2015), https://blogs.cisco.com/sp/the-history-and-future-of-internet-traffic [https://perma.cc/4QWC-SC8V].

4. *See* David Burg & Jesse H. Ausubel, *Moore's Law Revisited Through Intel Chip Density*, PLoS ONE, Aug. 18, 2021, at 1–2, https://doi.org/10.1371/journal.pone.0256245.

5. Lee Bell, *What Is Moore's Law? WIRED Explains the Theory That Defined the Tech Industry*, WIRED (Aug. 28, 2016, 12:00 PM), https://www.wired.co.uk/article/wired-explains-moores-law [https://perma.cc/4DW5-5JU6].

6. *But see id.* (attributing the rise of artificial intelligence to slowed capacity growth).

and artificial intelligences (AI) based on deep learning principles.[7] As a result of both these changes and the increasing primacy of the internet in day-to-day life, federal courts have struggled to balance the CDA's protections with the deluge of new types, forms, and producers of information content.[8]

As it stands, the CDA shields internet service providers not only from the information content posted by their users and other third parties, but also bestows protection on providers for decisions to remove or block information content.[9] However, there is a growing controversy about whether the CDA equally protects an internet service provider's affirmative decision to surface, or otherwise promote, third-party information content, rather than censor it.[10] The controversy is reducible to the foundational question of what constitutes information content creation or development: specifically, whether algorithmically tailored recommendations of third-party content based on detailed user data is itself new information content.

The current trend in the circuit courts is towards ever broader immunity for internet companies under Section 230 of the CDA premised on those companies' use of increasingly sophisticated technology.[11] This sweeping immunity for companies employing algorithmic processes is problematic because it enables those companies to escape liability for the harms the use of their technology causes and disincentivizes them from curbing or reducing future harms. Further, the way courts are treating machine learning algorithms under Section 230 has troubling implications for how they will treat such technologies deployed in an offline, or non-public, context.[12] Specifically, the obtaining interpretation of Section 230 implicates how courts will deal with issues of intent and causation when deep learning algorithms, through complexity and obscurity, break the

---

7. *See* Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age*, PEW RSCH. CTR. (Feb. 8, 2017), https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2017/02/PI_2017.02.08_Algorithms_FINAL.pdf [https://perma.cc/UWN5-76UQ].

8. *See* discussion *infra* Part II.

9. 47 U.S.C. § 230(c).

10. *See, e.g.*, Brief of Amicus Curiae Artificial Intelligence Law & Policy Institute in Support of Rehearing En Banc at 4–10, *Gonzalez v. Google LLC*, 2 F.4th 871 (9th Cir. 2021) (No. 18-16700) (arguing that surfacing algorithms should be treated differently than censoring algorithms under Section 230).

11. *See, e.g.*, *Gonzalez v. Google LLC*, 2 F.4th 871, 880, 897 (9th Cir. 2021); *Force v. Facebook, Inc.*, 934 F.3d 53, 57, 67 (2d Cir. 2019); *Marshall's Locksmith Serv. Inc. v. Google LLC*, 925 F.3d 1263, 1265, 1269 (D.C. Cir. 2019).

12. *See, e.g.*, DOROTHY J. GLANCY, ROBERT W. PETERSON & KYLE F. GRAHAM, A LOOK AT THE LEGAL ENVIRONMENT FOR DRIVERLESS VEHICLES 15–16, 34–35, 39 (2016); Clifford Law, *Driverless Car Accidents–Who's at Fault?*, NAT'L L. REV. (June 25, 2021), https://www.natlawreview.com/article/driverless-car-accidents-who-s-fault [https://perma.cc/C4D3-T6KW].

natural flow of both.[13] On the internet, Section 230 of the CDA, as currently constructed, fails to properly apportion liability to the developers and deployers of those deep learning algorithms.

Congress enacted the CDA in 1996 in an attempt to curb the growing problem of obscenity and other offensive content on the internet through the encouragement of self-censorship by internet service providers.[14] The CDA was meant to address the problem of offensive content in a way that did not stifle the free growth and exchange of the internet.[15] After its passage, the CDA was interpreted by the federal courts to be a broad grant of immunity to internet service providers hosting third-party content.[16] As part of the interpretative framework, many courts adopted the so-called "material contribution" test for determining when a person or entity is an "information content provider" and therefore subject to liability.[17] Under a component of this theory—"neutral tool" analysis—any algorithm not designed with unlawful intent will pass muster as not materially contributing to the information content.[18]

In applying the "material contribution" test, the courts have incorrectly concluded that advanced deep learning recommendation algorithms are "neutral tools" operating on "voluntary inputs" and thus are not information content providers.[19] By adopting this excessively crimped interpretation of "information content provider" under Section 230, courts have erroneously granted immunity to internet service providers that create or develop information through machine learning algorithms. Under a new "discretionary" test posited in this Comment,[20] "information content provider" assumes its proper scope, limiting Section 230 application only to information that is truly provided by third parties with no discretion in creation or development by an internet service provider. This discretionary test properly determines deep learning content recommendation algorithms to be information content providers.

The strongest criticism of interpreting "information content provider" under the CDA to include machine learning algorithms is that it could lead internet service providers to limit the utility of their services

---

13. *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 897 (2018).

14. *See* 47 U.S.C. § 230(b).

15. *Id.*

16. *See, e.g.*, *Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119, 1123 (9th Cir. 2003).

17. *Gonzalez v. Google, LLC*, 2 F.4th 871, 892 (9th Cir. 2021) (listing circuits that have adopted a material contribution standard).

18. *Id.*

19. *See* discussion *infra* Part II.

20. *See* discussion *infra* Part III.

and make the internet a less free-flowing and dynamic place.[21] The argument is that Congress intended the CDA to offer broad immunity to internet service providers to encourage the growth and vitality of the internet.[22] To address this criticism, this Comment will emphasize that narrowing Section 230, so as not to automatically immunize algorithms, by no means ensures liability for internet service providers. In fact, liability will continue to be limited by the natural difficulties deep learning algorithms and AI pose regarding intent, knowledge, proximate cause, and possibly free speech.[23] Thus, while internet companies will have a greater incentive to take proactive steps to curb the most harmful aspects of their platforms, they will not face enough risk to shutter or unduly limit their (lucrative) operations. Thus, the construction of "information content provider" should be accorded its full and fair scope,[24] not arbitrarily limited by the nature of the algorithm creating or developing the information.

This Comment argues that deep learning algorithms can be information content providers under Section 230 and ought to be treated as such by scuttling neutral tool analysis in favor of a more workable discretionary test. Part I provides necessary technical and legal background information for a thorough understanding of the application of Section 230 immunity to machine learning algorithms on the web. Part II explains the *Gonzalez v. Google LLC*[25] court's reasoning concerning neutral tools, voluntary inputs, and content surfacing algorithms before walking through a hypothetical applying the framework to a landmark case. Part III proposes an alternative framework—the discretionary test—for evaluating whether information content has been created or developed. The discretionary test determines whether an entity has the created or developed information content by analyzing if, given a set of user inputs, the outputs are discretionary based upon a choice of the entity's values, judgments, or inferences. Finally, Part IV discusses other barriers and defenses to liability that internet service providers should have if Section 230 immunity is pared back to its proper scope.

## I. THE COMMUNICATIONS DECENCY ACT AND MODERN TECHNOLOGY

This Part first discusses the passage of the CDA, including the text, purpose, and judicial construction of Section 230. Sections I.C and I.D outline the nature and capabilities of algorithms as well as the scope of data collection on the internet. Finally, Section I.E lays out the

21. *See Carafano*, 339 F.3d at 1123.
22. *Id.* at 1123–24.
23. *See* discussion *infra* Part IV.
24. *Cf.* ANTONIN SCALIA & BRYAN A. GARNER, READING LAW 101 (2012).
25. 2 F.4th 871 (9th Cir. 2021).

background facts and procedural posture of *Gonzalez v. Google*, which will provide a case study of the hardening consensus in the circuit courts for the application of neutral tool analysis to machine learning algorithms under Section 230.

### *A. Passage of the Communications Decency Act*

Congress enacted the CDA, the short name for Title V of the Telecommunications Act of 1996,[26] when the "world wide web" was still in its infancy.[27] The Telecommunications Act itself was not originally concerned with the internet at all but rather cable and broadcast services, with the provisions addressing the internet only proposed as floor amendments to the bill.[28] The Senate and House ultimately introduced and passed two such amendments, one from each chamber, as part of a unified CDA: Sections 223 and 230.[29] Both sections sought to address the evil of obscenity and harassment on the internet, particularly for children,[30] though through different means. Section 223 criminalized the transmission of obscene material or harassing communications over the internet, authorizing government agencies to monitor online traffic.[31] However, this provision was swiftly struck down by the Supreme Court in *Reno v. ACLU*,[32] finding that the statute violated the First Amendment's free speech protections.[33]

Section 230 took a different tack than that offered in Section 223 by empowering private actors to self-screen and self-regulate obscene material on the internet.[34] According to the legislative record, the amendment that became Section 230 was a direct reaction to state court cases disincentivizing self-regulation by internet service providers.[35] For example, in *Stratton Oakmont, Inc. v. Prodigy Service Co.*,[36] a New York trial court held a website host, Prodigy, liable for the defamatory

---

26. Telecommunications Act of 1996, Pub. L. No. 104-104, tit. 5, 110 Stat. 133–43 (codified as amended at 47 U.S.C. § 230).

27. *See World Wide Web Timeline*, PEW RSCH. CTR. (Mar. 11, 2014), https://www.pewresearch.org/internet/2014/03/11/world-wide-web-timeline [https://perma.cc/5DL3-VE7P].

28. *Force v. Facebook, Inc.*, 934 F.3d 53, 77–78 (2d Cir. 2019) (Katzman, C.J., concurring in part and dissenting in part).

29. *See id.* at 78–79.

30. 142 CONG. REC. 15503 (1995) (statement of Sen. J. James Exon).

31. Telecommunications Act of 1996, Pub. L. No. 104-104, § 502, 110 Stat. 133–36 (codified as amended at 47 U.S.C. § 223).

32. 521 U.S. 844 (1997).

33. *Id.* at 849.

34. *See* 141 CONG. REC. 22045 (1995) (statement of Rep. Charles Christopher Cox).

35. *Id.*

36. No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

statements its third-party users posted to its site.[37] The court found that the website—by employing automated screening tools for offensive language, promulgating content guidelines, and manually removing certain posts—was a "publisher" and therefore responsible for the content.[38] The court contrasted the "editorial control" exercised by Prodigy with the utter lack of screening present in *Cubby, Inc. v. CompuServe, Inc.*[39] In *Cubby*, the District Court for the Southern District of New York held that CompuServe was not liable for storing and making available a defamatory article in its database.[40] Because CompuServe did not review or censor the content added to its database, it did not act as a publisher but rather only as a distributor and a type of "for-profit library" or public repository.[41]

Thus, at the time Congress enacted the CDA, the internet was being regulated under a piecemeal liability scheme that granted immunity to websites that did not exercise any content restrictions but denied immunity to websites that implemented imperfect content controls. Section 230 was meant to address the perverse incentives this scheme gave internet companies to monitor—or rather not to monitor—obscene material on their websites.[42] Addressing Section 230's intent to protect kids from online obscenity on the House floor, sponsor Representative Jim Cox bemoaned the fact that "[i]ronically, the existing legal system provides a massive disincentive for the people who might best help us control the Internet to do so."[43] To rectify that disincentive, Section 230 was to "protect . . . online service providers, anyone who provides a front end to the Internet, . . . who takes steps to screen indecency and offensive material for their customers. It will protect them from taking on liability . . . ."[44] Further, the sponsors of Section 230 were concerned that the Senate's proposed Section 223, which gave the Federal Communications Commission power to regulate the web, threatened a free and open internet.[45] With the Supreme Court holding Section 223

37. *Id.* at *3–4.

38. *Id.* at *4.

39. *Id.* at *4–5 (citing *Cubby, Inc. v. CompuServe Inc.*, 776 F. Supp. 135, 140 (S.D.N.Y 1991)).

40. *Cubby, Inc.*, 776 F. Supp. at 137.

41. *Id.* at 140.

42. *See* 141 CONG. REC. 22045 (1995).

43. *Id.*

44. *Id.*

45. *Id.* ("[Section 230] will establish as the policy of the United States that we do not wish to have content regulation by the Federal Government of what is on the Internet, that we do not wish to have a Federal Computer Commission with an army of bureaucrats regulating the Internet because frankly the Internet has grown up to be what it is without that kind of help from the Government.").

unconstitutional,[46] Section 230—passed unchanged by Congress—became the sole source of obscene content regulation for the budding internet.[47]

### *B. Section 230's Text, Purpose, and Judicial Construction*

The statutory text and purpose of Section 230 express the Congressional intent to incentivize internet service providers to screen offensive content without fear of liability, not to grant blanket immunity to the use of certain types of algorithmic processes. Specifically, Section 230, entitled "Protection for private blocking and screening of offensive material," makes several factual findings and policy statements in its opening paragraphs.[48] The findings largely extoll the untold possibilities of the internet: from the democratization of information to user control.[49] The policy statements elaborate on the dual purposes of the CDA as its sponsors articulated in the Congressional record. Section 230 is "to promote the continued development of the Internet" and "preserve the vibrant and competitive free market . . . for the Internet" while "encourag[ing] the development of technologies [for] user control over what information is received" and "remov[ing] disincentives for . . . blocking and filtering technologies that empower parents to restrict their children's access" to offensive material.[50]

The operational core of the CDA is Section 230(c), entitled "Protection for 'Good Samaritan' blocking and screening of offensive material."[51] The provision seeks to distinguish between internet service providers and information content providers, consisting of two main components: the civil liability portion and the "treatment of publisher or speaker" portion.[52] The civil liability portion immunizes both providers and users of an interactive computer service against actions taken to restrict access to material deemed offensive or from the provision of the technical means to enable or facilitate such restrictions.[53] The "treatment of publisher or speaker" portion, which is the basis for the broad immunity currently given by the courts, states that "[n]o provider or user of an interactive computer service shall be treated as the publisher or

46. *Reno v. ACLU*, 521 U.S. 844, 849 (1997).

47. *Force v. Facebook, Inc.*, 934 F.3d 53, 79 (2d Cir. 2019) (Katzman, C.J., concurring in part and dissenting in part).

48. 47 U.S.C. § 230.

49. § 230(a).

50. § 230(b).

51. § 230(c).

52. *Id.*

53. § 230(c)(2).

speaker of any information provided by another information content provider."[54]

An "internet service provider," rarely at issue in Section 230 litigation, is defined as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server . . . ."[55] More importantly for the purposes of this Comment—and the source of contention in much Section 230 litigation—is the role of the "information content provider," statutorily defined 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."[56] The statute is rounded out by Section 230(e), which provides exceptions to immunity for various other laws, including criminal and intellectual property laws.[57] The only two amendments to the original Section 230 text are both in Subsection (e), addressing Section 230's inapplicability to federal criminal and sex trafficking laws.[58]

Regarding procedural application, there is no agreement, or much judicial analysis, on when Section 230(c) protection may be raised.[59] The majority of courts that have addressed the question have regarded Section 230(c) as providing immunity to defendants.[60] Other courts have treated it as an affirmative defense, which means a defendant may not raise it on a motion to dismiss unless the complaint itself establishes the defense.[61] Thus, whether a court characterizes Section 230's protections as immunity or an affirmative defense has ramifications for a plaintiff's ability to engage in discovery. Discovery may prove particularly important for Section 230 cases involving machine learning and artificial intelligence, since the exact nature of the algorithms employed, as well

---

54. § 230(c)(1). There may be a circuit split concerning Section 230(c)(1) between an immunity approach and a "definitional" approach, the latter of which does not confer blanket immunity on internet service providers for third-party content. *See* Ellen Smith Yost, *Social Support for Terrorists: Facebook's "Friend Suggestion" Algorithm, Section 230 Immunity, Material Support for Terrorists, and the First Amendment*, 37 SANTA CLARA HIGH TECH. L.J. 301, 314 (2021).

55. § 230(f)(2).

56. § 230(f)(3).

57. § 230(e).

58. *See* Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115-164, 132 Stat. 1253–55 (2018).

59. *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 482–83 (2010).

60. *Id.* at 483 (first citing *Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119, 1125 (9th Cir. 2003); and then citing *Ben Ezra, Weinstein, & Co. v. Am. Online, Inc.*, 206 F.3d 980, 983 (10th Cir. 2000)).

61. *See, e.g.*, *Gonzalez v. Google LLC*, 2 F.4th 871, 890 n.8 (9th Cir. 2021); *see also* Ardia, *supra* note 59, at 483 n.370.

as the source and form of the relevant information, may affect the analysis of information content development and immunity.[62] However, given that courts have ruled on Section 230 before full discovery in 92.4 percent of cases, plaintiffs currently have little opportunity for developing such facts.[63]

The dominant analytical framework for assessing algorithms vis-à-vis information content providers under Section 230 is "neutral tool" analysis, part of the "material contribution test."[64] First promulgated in *Fair Housing Council v. Roommates.com, LLC*,[65] at least six circuit courts have adopted—explicitly or implicitly—the material contribution test's neutral tool analysis.[66] Under the material contribution test, in order to be considered an "information content provider" of the content at issue, the internet service provider must contribute "materially to the alleged illegality of the conduct."[67] Neutral tool analysis looks to the nature of the algorithm and a user's inputs when determining the material contribution of algorithmic processes.[68]

In *Roommates.com*, the defendant operated a website that matched users leasing spare rooms with potential renters.[69] When a user registered for the site, they were required to disclose their gender, sexual orientation, and familial status as well as their preferences for those characteristics in others.[70] Plaintiffs sued, alleging violations of the Fair Housing Act.[71] The Court of Appeals for the Ninth Circuit held that the claims were not barred by Section 230.[72] In reaching its conclusion, the court emphasized that the site was "*designed* to steer users based on discriminatory criteria" and "*force[d]* subscribers to disclose" their personal characteristics as conditions of use.[73] Thus, the site's registration questions elicited, and therefore developed, the content on which the illegal discrimination was based.[74] The court further

62. *See* discussion *infra* Section II.A.

63. *See* Ardia, *supra* note 59, at 483.

64. *See Gonzalez*, 2 F.4th at 892–93.

65. 521 F.3d 1157 (9th Cir. 2008) (en banc).

66. *See, e.g.*, *id.* at 1168–69; *Force v. Facebook, Inc.*, 934 F.3d 53, 68 (2d Cir. 2019); *Jones v. Dirty World Ent. Recording LLC*, 755 F.3d 398, 411–16 (6th Cir. 2014); *Marshall's Locksmith Serv. Inc. v. Google, LLC*, 925 F.3d 1263, 1270–71 (D.C. Cir. 2019); *Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc.*, 591 F.3d 250, 257 (4th Cir. 2009); *FTC v. Accusearch Inc.*, 570 F.3d 1187, 1200 (10th Cir. 2009); *Gonzalez*, 2 F.4th at 892–93.

67. *See Roommates.com*, 521 F.3d at 1168.

68. *See id.* at 1169, 1172.

69. *Id.* at 1161.

70. *Id.*

71. *Id.* at 1162.

72. *Id.* at 1165.

73. *Id.* at 1167 (emphases added).

74. *Id.* at 1166–67.

distinguished Roommates.com's search function, which was designed to promote unlawful discrimination, from "ordinary search engines," which "do not use unlawful criteria to limit the scope of searches conducted on them [and are not] *designed* to achieve illegal ends."[75] The court regarded these ordinary search engines as "neutral tools" which merely performed searches based on "user-defined criteria" or "voluntary inputs" which may or may not be unlawful.[76]

As an example of the neutral tools, in addition to an "ordinary search engine," the court reexamined the facts of *Carafano v. Metrosplash.com, Inc.*,[77] in which a third-party user posted a fake dating profile of an actress containing libelous content on the defendant's dating site.[78] The *Carafano* court, applying an expansive reading of Section 230, found the website immune even though the site prompted the user for particular inputs, similar to *Roommates.com*.[79] The *Roommates.com* court's posthumous analysis determined that, though the holding was correct, the reasoning was too broad.[80] In retrospect, the dating site was immune because it merely provided a "neutral tool" by which users, without encouragement, could upload content and match based on those voluntary inputs.[81] That is, neither the website's match-making function nor the user profile posting function contributed to the alleged unlawfulness because they were neutral tools.[82] Thus, there is no material contribution when an internet service provider employs a "'neutral tool' operating on 'voluntary inputs.'"[83]

### *C. Artificial Intelligence and Machine Learning Algorithms*

To appreciate fully the application of Section 230 immunity to algorithmic recommendation systems, it is necessary to have a general baseline understanding of algorithms. The Computer Science 101 example of an algorithm is a recipe: a series of steps (*e.g.*, mix flour, eggs, milk) to produce a particular outcome (*e.g.*, a cake). These simple algorithms involve a series of discrete steps with deterministic outcomes given the same inputs.[84] Put another way, the computer programmer develops and codes rules instructing a computer on how to solve a

75. *Id.* at 1167 (emphasis added).
76. *Id.* at 1169, 1172.
77. 339 F.3d 1119 (9th Cir. 2003).
78. *Roommates.com*, 521 F.3d at 1171 (citing *Carafano*, 339 F.3d at 1121).
79. *Carafano*, 339 F.3d at 1124.
80. *Roommates.com*, 521 F.3d at 1171.
81. *Id.* at 1171–72.
82. *See id.*
83. *See Kimzey v. Yelp! Inc.*, 836 F.3d 1263, 1270 (9th Cir. 2016).
84. *See* Bathaee, *supra* note 13, at 898.

discrete problem.[85] In that scenario, the programmer needs to anticipate all possible inputs and account for them in their instructions. If the program gets an unrecognized input that the given rules do not account for, the program will fail. For example, a rule-based program to play chess might require the program to walk through every permutation of possible moves and choose the best outcome with a certain scoring formula.[86]

Machine learning sits in contrast to such rule-based programming.[87] In machine learning, instead of programming the computer to solve the problem directly, the programmer codes the program to learn to solve the problem.[88] To return to the chess example, a machine learning algorithm might, after being fed a host of historical chess games (often called training data),[89] determine its own scoring criteria and method of play.[90] The training data set determines the decisions that the algorithm makes.[91] Thus, though the programmer imparts no bias, a machine learning algorithm developed based on biased data will replicate that bias in its decisions. Amazon recently discovered this lurking peril of machine learning algorithms when its hiring AI ended up replicating the gender discrimination in the historical employment data on which the AI was trained.[92]

However, a machine learning algorithm need not be limited to training sets, which programmers often curate for their simplicity or breadth. Continual learning AI update their self-created "rules" with each new game (in our chess example) or other piece of information they encounter after being developed and deployed based on the training data set.[93] This means that programmers can cede control over what data the machine is being trained on such that the AI becomes even more likely

---

85. *See id.*

86. *Id.*

87. *Id.*

88. *Id.* at 898–99.

89. *See id.* at 900.

90. *See id.* at 898–900. *Accord* David Gershgorn, *Artificial Intelligence Is Taking Computer Chess Beyond Brute Force*, POPULAR SCI. (Sept. 16, 2015, 9:07 PM), https://www.popsci.com/artificial-intelligence-takes-chess-beyond-brute-force [https://perma.cc/SQL9-R29J].

91. *See* Jason Tanz, *Soon We Won't Program Computers. We'll Train Them Like Dogs*, WIRED (May 17, 2016, 6:50 AM), https://www.wired.com/2016/05/the-end-of-code [https://perma.cc/SM4N-LWLR].

92. Jeffery Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 6:04 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G [https://perma.cc/P99M-HQXL].

93. *See, e.g.*, Raia Hadsell, Dushyant Rao, Andrei A. Rusu & Razvan Pascanu, *Embracing Change: Continual Learning in Deep Neural Networks*, 24 TRENDS IN COGNITIVE SCIS. 1028, 1028 (2020).

to replicate the hidden biases in our society without its developers' knowledge.

One of the most common types of machine learning algorithms being deployed by many software companies such as Google and Facebook[94] is deep neural networks.[95] As the name suggests, deep neural networks seek to replicate the organic learning processes of the brain.[96] In these algorithms, artificial neurons are interconnected in massive webs of logical and categorical relations.[97] Like with the human brain, no singular neuron or node "encodes a distinct part of the decision-making process."[98] Instead, the connections and patterns formed throughout the entire network allow the machine to make decisions.[99] While these connections allow the machine to make decisions, asking how the system arrived at any particular outcome is like attempting to describe how someone ordered off a drive-thru menu by reference to an MRI scan.[100]

The decision-making process for a deep neural network is often completely opaque, even to the programmers who created it.[101] In fact, not only is how the AI arrived at a decision impossible to know, but what data was outcome determinative or what variables played a role are equally impenetrable.[102] There are a couple of reasons for this impenetrability. The complexity of the algorithmic structure, consisting of the perhaps billions of independent neurons, defies human imagination or ability to comprehend.[103] A programmer who opens a deep neural network to look at a particular moment in time will see nothing but billions or trillions of data points and "a massive, multilayer set of calculus problems."[104]

The lack of transparency in an AI's decision-making processes poses a problem for many aspects of the law. For example, in Amazon's hiring algorithm mentioned above, the AI was making decisions based on unlawful criteria.[105] However, even if the AI's programmers went back in to code a set of hard blocks on using gender as a criterion in its analysis, a deep neural network is capable of finding and employing

94. Tanz, *supra* note 91.
95. *See* Bathaee, *supra* note 13, at 902–03.
96. *Id.* at 901.
97. *Id.* at 902.
98. *Id.*
99. *Id.*
100. *See id.* at 902–03.
101. *Id.* at 906.
102. *Id.*
103. Tanz, *supra* note 91.
104. *Id.*
105. *See* Dastin, *supra* note 92.

decision-making protocols that simply rely on proxies for gender.[106] Thus, there would be no way of knowing whether the algorithm is making its hiring recommendations based solely on lawful characteristics.[107]

### *D. Internet Data Collection*

In the age of big data, information collection on the internet is ubiquitous.[108] First, many websites require users to provide some personal information to register to use the website, such as name, email address, physical address, telephone number, etc.[109] But most data collection by websites is not done transparently to the user.[110] While on a website, the site may be tracking a user's every click, hover, keystroke, or other user interaction on each page.[111] Sites may automatically collect the user's unique IP (internet protocol) address upon visiting a website.[112] Further, nearly all websites, when a user loads a page, will place a cookie on the user's computer that acts as a unique identifier and allows the website to track the particular user's comings and goings on the website.[113] Even when a user has not visited a website, third-party websites may set cookies on the user's computer if another website the user visits loads an image or other data from the third-party site.[114] This third-party cookie setting has been exploited in the form of so-called "pixel tags" or "web beacons," which are small, imperceptible images loaded on millions of websites, serving no other purpose than to enable

---

106. *See* Bathaee, *supra* note 13, at 907; *see also* Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1273–76 (2020).

107. *See* Bathaee, *supra* note 13, at 907; see also Prince & Schwarcz, *supra* note 106, at 1264.

108. Dave Davies, *How Tech Companies Track Your Every Move and Put Your Data Up for Sale*, NPR (July 31, 2019, 1:29 PM), https://www.npr.org/2019/07/31/746878763/how-tech-companies-track-your-every-move-and-put-your-data-up-for-sale [https://perma.cc/EJ4M-GSQZ].

109. *See* Abdelmounaam Rezgui, Athman Bouguettaya & Mohamed Y. Eltoweissy, *Privacy on the Web: Facts, Challenges, and Solutions*, IEEE SEC. & PRIV., Nov.–Dec. 2003, at 43.

110. John A. Rothchild, *Against Notice and Choice: The Manifest Failure of the Proceduralist Paradigm to Protect Privacy Online (or Anywhere Else)*, 66 CLEV. ST. L. REV. 559, 567–68 (2018).

111. *See* Benjamin Mangold, *Google Analytics 4 Event Tracking: Your Complete Guide*, LOVESDATA, https://www.lovesdata.com/blog/google-analytics-4-events [https://perma.cc/Y3ME-QAS6] (last visited Oct. 30, 2022).

112. Ivan Dimov, *Means and Methods of Web Tracking: Its Effects on Privacy and Ways to Avoid Getting Tracked*, INFOSEC (July 23, 2013), https://resources.infosecinstitute.com/topic/means-and-methods-of-web-tracking-its-effects-on-privacy-and-ways-to-avoid-getting-tracked [https://perma.cc/TB8H-FY8K].

113. Rothchild, *supra* note 110, at 568–69, 577.

114. Dimov, *supra* note 112.

cookies to be set and thereby track a user around the web.[115] These various tracking methods "allow much of a user's clickstream data, generated through visits to many websites, to be assembled into a single dossier that is associated with that individual."[116]

Google, as the market leader in online advertising,[117] has arrangements with millions of websites, allowing it to place cookies over large swaths of the internet and assemble dossiers on millions of internet users.[118] Such massive data collection allows Google "to better identify users of the Web and track them as they engage in various online activities."[119] Moreover, the information collected is not limited to what we do online. For example, Google knows when someone who has previously clicked on one of its ads later buys the product in a brick-and-mortar store, thanks to data sharing partnerships with credit card companies.[120] Facebook has made similar deals that allow it to target advertisements based on its users' recent purchases at physical stores.[121] Though there are occasionally ways to opt out or otherwise obscure the voluminous data collected by internet companies, the typical user—often unaware the data is even being collected—is unlikely to know how to go about maximizing their privacy.[122] The scope and breadth of data collection on internet users, even people not logged in to any particular site, is massive and allows internet companies to develop deep profiles on individuals without users' knowledge or (knowing) consent.[123]

### *E.* Gonzalez v. Google LLC *Facts and Procedural Posture*

Circuit courts have continually applied and revised neutral tool analysis to further restrict when an internet service provider may be liable

---

115. Rothchild, *supra* note 110, at 568 n.39.

116. *Id.* at 568.

117. Megan Graham & Jennifer Elias, *How Google's $150 Billion Advertising Business Works*, CNBC, https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html [https://perma.cc/F9RJ-KCTG] (Oct. 13, 2021, 12:52 PM).

118. *See* Rothchild, *supra* note 110, at 568–69.

119. *Id.* at 569.

120. Elizabeth Dwoskin & Craig Timberg, *Google Now Knows When Its Users Go to the Store and Buy Stuff*, WASH. POST: THE SWITCH (May 23, 2017, 8:00 PM), https://www.washingtonpost.com/news/the-switch/wp/2017/05/23/google-now-knows-when-you-are-at-a-cash-register-and-how-much-you-are-spending [https://perma.cc/3HA3-8A8Q]; *see also* Graham & Elias, *supra* note 117.

121. Aaron Holmes, *This Is How Facebook Learns What You Buy at Physical Stores in Order to Show You Relevant Ads — and How to Opt Out*, BUS. INSIDER (Dec. 11, 2019, 2:54 PM), https://www.businessinsider.com/facebook-learns-what-you-buy-at-physical-stores-ads-explained-2019-12 [https://perma.cc/2A7X-JSGU].

122. *See* Dimov, *supra* note 112.

123. *See* Rezgui, Bouguettaya & Eltoweissy, *supra* note 109, at 43.

as the creator or developer information, most recently with the Ninth Circuit in *Gonzalez v. Google LLC*.[124] In *Gonzalez*, the plaintiff brought claims pursuant to the Anti-Terrorism Act's (ATA) civil remedies provision,[125] alleging Google, through its subsidiary YouTube, provided material support to the terrorist group ISIS (the Islamic State of Iraq and Syria) by, among other things, recommending ISIS videos to users.[126] The case consolidated three separate appeals, addressing liability under the ATA.[127] However, only the facts and holding concerning the Gonzalez family are relevant for Section 230 immunity. The *Gonzalez* plaintiffs were the family of a U.S. citizen who was studying abroad in Paris when she was killed on November 13, 2015, during the ISIS terrorist attacks around the city.[128]

The *Gonzalez* complaint alleged that the perpetrators of those attacks were recruited, trained, and radicalized by propaganda videos that ISIS, through one of its many well-polished media arms, posted to YouTube.[129] The complaint further alleged that ISIS, through YouTube and other social media, has recruited more than 30,000 volunteers to its cause.[130] ISIS managed to accomplish those staggering recruitment numbers allegedly due to Google's platform, which is "comprised of highly advanced software, algorithms, . . . computer applications and more."[131] The "sophisticated technological capabilities" of Google's platform allegedly had an enormous impact on ISIS's success in "recruiting, indoctrination, training, conducting terrorist operations, and engaging in psychological warfare."[132]

Among those alleged sophisticated technologies, Google's recommendation system and targeted advertising scheme are specifically noted.[133] Google's machine learning algorithms recommend "content to users based upon the content and what is known about the viewer," including what videos the users had viewed in the past.[134] Those algorithms match users to both videos and other users according to the content of the videos as well as the user's account information and

---

124. 2 F.4th 871 (9th Cir. 2021).

125. *Id.* at 880; 18 U.S.C. § 2333.

126. *Gonzalez*, 2 F.4th at 881–82.

127. *Id.* at 880. The Second Circuit recently dealt with the same issue and arrived at the same conclusion with the same reasoning, so *Gonzalez* will serve as a case study for the category. *See Force v. Facebook, Inc.*, 934 F.3d 53, 61–62, 68–71 (2019).

128. *Gonzalez*, 2 F.4th at 879–81.

129. Third Amended Complaint at 40–46, 57, 65, 68, *Gonzalez v. Google, Inc.*, 335 F. Supp. 3d 1156 (N.D. Cal. 2018) (No. 16-cv-03282).

130. *Id.* at 45.

131. *Id.* at 31.

132. *Id.* at 52.

133. *Id.* at 96, 99.

134. *Id.* at 99.

characteristics.[135] Further, the recommended videos are loaded and play automatically when a different video ends.[136] The complaint alleged ISIS videos had been recommended to YouTube users frequently.[137]

The complaint also highlighted the targeted advertising system Google employs on YouTube.[138] The system targeted advertisements to users based on the content of the video and information about the user.[139] Because targeted advertising draws complex connections between the content of the advertisement, the content of the video, and the information about the user, Google was able to extract a premium fee from advertisers for this service.[140] Additionally, a percentage of the advertising revenue generated from such algorithms went to the account that posted the video: in this case, ISIS recruiter accounts.[141]

To access certain features on YouTube, a user must register for a Google account.[142] The registration process required users to disclose information.[143] In addition to the identifying information—including name, telephone number, and email address—a user allegedly must input to complete registration, Google automatically collected IP address, geographical information, and other data that enables Google to determine a user's other, undisclosed, Google accounts.[144] Beyond the registration process, the complaint alleged that Google also collected information about a user's site usage such as which videos they have posted, viewed, liked, commented on, subscribed to, and shared.[145]

The District Court for the Northern District of California granted Google's motion to dismiss plaintiffs' complaint.[146] In the dismissal order, the court concluded that the content recommendation claims were barred by Section 230 of the CDA.[147] Specifically referring to the material contribution's "neutral tool" analysis developed in *Roommates.com*, the court found that "[a]s with Google's targeted ad algorithm, there is no indication that its content recommendation tool is anything other than content neutral."[148] The court reasoned that the "use

135. *Id.* at 101.
136. *Id.* at 101–02.
137. *Id.* at 99.
138. *Id.* at 96.
139. *Id.*
140. *Id.* at 95–96, 100.
141. *Id.* at 95–96.
142. *Id.* at 32.
143. *Id.*
144. *Id.* at 32–33.
145. *Id.* at 100–01.
146. *Gonzalez v. Google, Inc.*, 335 F. Supp. 3d 1156, 1160 (N.D. Cal. 2018).
147. *Id.* at 1172–74.
148. *Id.* at 1173. This is a problematic statement because in a motion to dismiss the court must accept as true all factual matter in a complaint, *Ashcroft v. Iqbal*, 556

of an algorithm that aggregates user and video data to make content recommendations across YouTube, whether the recommended content is an ISIS video or a cat video, does not turn Google into an 'information content provider' with respect to the videos themselves."[149]

On appeal, the Ninth Circuit affirmed the lower courts' decision to dismiss.[150] The panel held that Google was immunized from liability under Section 230, finding that Google did not "materially contribute" to its own recommendations.[151] In so holding, the court undertook a lengthy examination and restatement of the Section 230 precedents regarding the material contribution test's neutral tools analysis.[152] At the outset, the court dismissed the plaintiffs' argument that consideration of Section 230 at the motion to dismiss stage is improper, finding the "allegations in the complaint suffic[ient] to establish that [affirmative defense]."[153] By refining the material contribution test into an ever narrower reading of "information content provider," the court continued to expand Section 230 immunity for internet service providers that create or develop information through algorithmic processes.

## II. NEUTRAL TOOL ANALYSIS UNDER *GONZALEZ V. GOOGLE LLC* FAILS TO ACCOUNT FOR THE SOPHISTICATION OF MACHINE LEARNING ALGORITHMS IN SURFACING CONTENT

Neutral tool analysis, as expounded in *Gonzalez* and similar cases, has proven unworkable in the face of increasingly sophisticated algorithms. This Part analyzes how courts have responded to the development of machine learning algorithms and mass data collection in the context of the CDA and neutral tool analysis. Sections II.A and II.B discuss the *Gonzalez v. Google* court's reasoning regarding neutral tools, voluntary inputs, and content surfacing algorithms. Using the neutral tool framework thus established, Section II.C will walk through a hypothetical applying that framework to an algorithmic version of the *Roommates.com* facts to show how the analysis has undergone substantial mission creep since its inception. This hypothetical

---

U.S. 662, 678 (2009), and the complaint specifically alleged that the algorithm makes recommendations based on content, Third Amended Complaint, *supra* note 129, at 99. The court seems to be taking judicial notice that Google's matching algorithms are commonly understood to be "content neutral." *See* discussion *infra* Section II.A.

149. *Gonzalez*, 335 F. Supp. 3d at 1173.

150. *Gonzalez v. Google LLC*, 2 F.4th 871, 880 (9th Cir. 2021), *cert. granted,* No. 21-1333, 2022 WL 4651229 (U.S. Oct. 3, 2022), *and cert. granted, sub nom. Twitter, Inc. v. Taamneh,* No. 21-1496, 2022 WL 4651263 (U.S. Oct. 3, 2022).

151. *Id.* at 896.

152. *Id.* at 891–97.

153. *Id.* at 890 n.8 (quoting *Jones v. Bock*, 549 U.S. 199, 215 (2007)). Again, this raises issues under *Iqbal. See* discussion *supra* note 148.

demonstrates how the framework is unable to account for machine learning algorithms behaving as information content providers, thereby underscoring that neutral tool analysis ought to be repudiated in favor of the more nuanced discretionary test.

*Gonzalez v. Google* functions as an apt case study in how the courts' use of neutral tool analysis has failed to grapple with increasingly advanced nature of the algorithms employed by internet service providers. In *Gonzalez*, the Ninth Circuit affirmed the grant of Section 230 immunity to Google, holding that Google did not *materially contribute* to its own video recommendations.[154] The elements for immunity under the CDA are (1) the defendant is a provider or user of an interactive computer service (2) whom the plaintiff's claim treats as a publisher or speaker (3) of information provided by a third-party information content provider.[155] The court quickly dismissed the first prong, as Google's status as an interactive computer service was undisputed.[156] The court disposed of the second prong almost as quickly, finding that the claims attempted to treat Google as the publisher of ISIS's videos in order to impose liability.[157] However, the court never addressed who was being treated as the speaker or publisher of the recommendations themselves, foreshadowing the essential proposition that the recommendations are not themselves information content.[158] Google is, of course, the speaker and publisher of its own recommendations, but the court failed to distinguish between the videos and the recommendations of the videos in its brief analysis of the second prong.[159]

The third prong is where the court began its analysis in earnest.[160] Per statute, in order to be an "information content provider" one must

---

154. *Id.* at 880.

155. *Id.* at 891 (quoting *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1100–01 (9th Cir. 2009)); *see also Marshall's Locksmith Serv. v. Google, LLC*, 925 F.3d 1263, 1267–68 (D.C. Cir. 2019) (quoting *Klayman v. Zuckerberg*, 753 F.3d 1354, 1357 (D.C. Cir. 2014)).

156. *Id.* at 891.

157. *Id.*

158. *Id.*

159. *Id.* at 891–92.

160. That information needs to be "provided by" a third-party information content provider is never analyzed. For example, in *Kimzey v. Yelp! Inc.,* the court held that Yelp was immune under Section 230 from liability related to negative reviews of plaintiff's locksmith business hosted on the site. 836 F.3d 1263, 1270 (9th Cir. 2016). The complaint alleged that rather than the third-party itself posting on Yelp, the company had scraped the review from another website entirely and displayed it on its own page. *Id.* It is not clear why websites that scrape data from other websites should be immune under the CDA, since no third party is "providing" the content to the internet service provider. Merriam Webster's Dictionary defines "provide" as some variation of "to make something available," or "to supply something." *Provide*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/provide [https://perma.cc/YU2R-2PHH]

be "responsible, in whole or part, for the creation or development of information provided through the Internet . . . ."[161] The court quickly brushed past the defining rules of the material contribution test—a website creates or develops content when it "materially contribut[es] to its alleged unlawfulness"[162]—before jumping into its "neutral tool" and "voluntary input" exegesis.[163] The court ultimately concluded, relying on recent precedent, that Google did not create or develop the videos.[164] However, the court also erroneously concluded that Google's recommendations of the videos at issue were not in themselves information content separable from the videos, based on its recent decision in *Dyroff v. Ultimate Software Group*.[165]

In *Dyroff*, the defendant operated a social networking website (Experience Project) that allowed users to anonymously share experiences, post and answer questions, and interact with other users on different topics.[166] The website did not collect any identifying information when users registered in order to facilitate open, anonymous sharing.[167] Further, the topics of groups on the website were entirely created by users, which users could opt to join.[168] However, the website also created recommendations for groups a user might be interested in joining "based on the content of their posts and other attributes, using machine learning algorithms."[169] Because the topics and content of the website were

---

(last visited Nov. 8, 2022). While the website might be "making something available" by posting it publicly, they certainly are not supplying or giving it to the third-party site. The courts' reading is analogous to an employee stealing a pharmacist's inventory and later claiming to the police the pharmacist "provided" them with the drugs.

A reading requiring the active participation by third party information content providers would add a fourth element to general test for immunity under the Section 230. Namely, that the information content provider supplied the information to the internet service provider, rather than the internet service provider taking it unbidden from other sources. Of course, there might be an argument that third-party sites are indeed offering their information content to aggregation sites like Google search engine and Yelp. However, if that is the case, then they are using what is known as a "robot.txt" file, which expressly allows website scrapes like those employed by Google and Yelp to access their sites. Absent that file, or similarly broad terms of service, there is a real question of whether the information at issue in some Section 230 cases are being "provided by" anyone other than the internet service provider who scraped the data.

161. *Gonzalez*, 2 F.4th at 892 (quoting 47 U.S.C. § 230(f)(3)).

162. *Id.* at 892 (quoting *Fair Hous. Council San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167–68 (9th Cir. 2008)).

163. *Id.* at 893 (quoting *Roommates*, 521 F.3d at 1172).

164. *Id.* at 893–94.

165. 934 F.3d 1093 (9th Cir. 2019). *See Gonzalez*, 2 F.4th at 895.

166. *Dyroff*, 934 F.3d at 1094.

167. *Id.* at 1095.

168. *Id.* at 1094–95.

169. *Id.* at 1095.

unregulated, many forums and groups included discussion and facilitation of illegal activity, including drug use and sales.[170]

Plaintiff Dyroff was the mother of a former user of the Experience Project website.[171] The plaintiff's son, who had a history of addiction, had died after he posted in a heroin related group that he was looking to buy the drug and subsequently bought heroin laced with fentanyl from another user.[172] The plaintiff alleged, among other things, that the website steered her son toward the heroin group where he ultimately bought the deadly dose.[173] The plaintiff further alleged that the user-specific group recommendations generated by the defendant's machine learning algorithms constituted information content and that defendant was therefore an information content provider.[174]

The court rejected the plaintiff's recommendation claims, affirming the defendant's Section 230 immunity.[175] In analyzing the *Barnes* prongs, the court concluded that the recommendations were not information content but "content-neutral website functions" that were "meant to facilitate the communication and content of others" but "are not content in themselves."[176] The court characterized the plaintiff's arguments as trying to "plead around Section 230 immunity by framing these website features as content."[177] The court cited no authority, persuasive or otherwise, for its assertion that recommendations generated based on a user's "posts and other attributes" are not information content.[178]

In *Gonzalez*, the court expanded on the (largely absent) reasoning of *Dyroff* for why recommendations are not information content, seemingly concluding that no product of neutral tools operating on voluntary inputs can ever be information content.[179] While summarizing several different cases, the court failed to adequately define a neutral

---

170. *Id.* at 1094–95.

171. *Id.* at 1094.

172. *Id.* at 1095.

173. *Id.* at 1095, 1098.

174. *Id.* at 1096.

175. *Id.* The court also concluded that defendant owed the deceased no duty of care, because the website's "content-neutral function[] . . . did not create a risk of harm." *Id.* at 1100. The assertion here—that things not designed to be harmful cannot create a risk of harm—is absurd on its face but is especially absurd with regard to machine learning and artificial intelligence, where experts warn of their dangers. *See* Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC NEWS (Dec. 2, 2014), https://www.bbc.com/news/technology-30290540 [https://perma.cc/K4T5-XQPB].

176. *Dyroff*, 934 F.3d at 1097–98.

177. *Id.* at 1098.

178. *Id.* at 1097–98.

179. *Gonzalez v. Google LLC*, 2 F.4th 871, 895 (9th Cir. 2021), *cert. granted,* No. 21-1333, 2022 WL 4651229 (U.S. Oct. 3, 2022), *and cert. granted, sub nom. Twitter, Inc. v. Taamneh,* No. 21-1496, 2022 WL 4651263 (U.S. Oct. 3, 2022).

tool, instead using, and neglecting to differentiate between, various meanings. Further, the court expanded its understanding of "voluntary input,"[180] without directly addressing what makes something either "voluntary" or an "input." By employing this neutral tool analysis without accounting for the algorithms being deployed or the nature of the data being operated upon, as discussed in the following sections, the court erroneously concluded that Google's recommendation algorithm is not an information content provider.

### *A. Neutral Tools and the Non-Neutrality of Machine Learning Algorithms*

The *Gonzalez* court began its discussion of neutral tools by quoting precedent to the effect that a "website does not create or develop content when it merely provides a neutral means by which third parties can *post* information of their own independent choosing online."[181] However, a few lines later, discussing *Roommates.com*, the court stated, "a website is not transformed into a content creator or developer by virtue of supplying 'neutral tools' that *deliver* content in response to user inputs."[182] Shortly thereafter, discussing *Carafano v. Metrosplash.com*, the court stated that the tools in that case "were neutral because the website did not 'encourage the *posting* of defamatory content' by merely providing a means for users to publish the profiles they created."[183] Finally, the court quotes *Roommates.com* as observing "search engines are immune under [Section] 230 because they provide content in response to a user's queries 'with no direct encouragement to *perform* illegal searches or to *publish* illegal content.'"[184]

Thus, the court identified, albeit implicitly, several forms of supposedly neutral tools. The first are the tools websites deploy to allow users to upload and publish information content to the site. The second are tools websites deploy to allow users to surface the information content of others. A neutral publishing tool is a fairly straightforward proposition: it need only not encourage users to post unlawful content.[185] Put another way, a neutral publishing tool does not distinguish the content being published and might only discriminate against unlawful content by,

180. *Id.* at 893, 895.

181. *Id.* at 893 (emphasis added) (quoting *Kimzey v. Yelp! Inc.*, 836 F.3d 1263, 1270 (9th Cir. 2016)).

182. *Id.* (emphasis added) (quoting *Fair Hous. Council San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1171 (9th Cir. 2008).

183. *Id.* (emphasis added) (quoting *Roommates.com*, 521 F.3d at 1171).

184. *Id.* at 895 (emphases added) (quoting *Roommates.com*, 521 F.3d at 1175).

185. *Id.* at 893.

for example, screening out profanity.[186] If a publishing tool does not attempt to screen any content, then it certainly complies with the material contribution test.[187] However, if the internet service provider attempts to implement some form of screening tool before allowing a user to post content, it is possible for it "encourage" users to post unlawful content.[188]

In contrast to the relative simplicity of neutral publishing tools, it is not so clear what the court meant by neutral surfacing tools. Such a tool cannot be content-neutral in the same way that a publishing tool can be because, put simply, the purpose of the tool is to match the content of user inputs with the content of videos.[189] There are two distinct forms of surfacing tools. Take what the court referred to as a "traditional search engine:" a textbox on a screen that the user types a query into and then has results returned.[190] The first form is the textbox: a tool that does not encourage users to perform illegal searches—like a tool that does not encourage users to publish illegal content—is neutral.[191] The second form, however, is the tool that actually returns the results requested by the user.[192] The court did not distinguish between these forms of surfacing tools, which muddies the analysis and mistakenly implies that all algorithms are equivalent no matter their function.[193]

When referring to surfacing tools as neutral beyond the textbox, it seems the court meant the tool acts as a neutral intermediary between the user inputs and the content. That is, a neutral intermediary would not prefer some content over other content except in so far as it was a "better" match for the users' inputs.[194] YouTube video recommendations would fall into this latter surfacing tool category, though with the caveat, discussed *infra*,[195] that the user inputs are not express.

Of course, whether a machine learning algorithm that recommends videos to users is acting as a neutral intermediary is a question of fact. The court avoided this factual issue, improper for resolution on a motion to dismiss, by implicitly reading an element of intent into the text of the

---

186. It is somewhat ironic that the court has decided to call tools that confer CDA immunity "neutral" when the whole point of the CDA was to promote non-neutrality against obscene and offensive content. *See* discussion *supra* Section I.B.

187. Another irony.

188. Whether a screening tool functions to encourage the posting of certain content is a question of fact.

189. *See Gonzalez*, 2 F.4th at 896.

190. *See id.* at 895.

191. *See id.* at 893.

192. *See id.* at 895.

193. *See id.* at 894–96.

194. *See Gonzalez v. Google LLC*, 2 F.4th 871, 894 (9th Cir. 2021), *cert. granted,* No. 21-1333, 2022 WL 4651229 (U.S. Oct. 3, 2022), *and cert. granted, sub nom. Twitter, Inc. v. Taamneh,* No. 21-1496, 2022 WL 4651263 (U.S. Oct. 3, 2022).

195. *See* discussion *infra* Section II.B.

CDA.[196] The complaint, the court opines, "is devoid of allegations that Google specifically *targeted* ISIS content, or *designed* its website to encourage videos that further the terrorist group's mission."[197] Rather, the complaint only succeeds in alleging that "Google provided a neutral platform that did not specify or prompt the type of content to be submitted, nor determine particular types of content its algorithms would promote."[198]

The court's neutral tool analysis failed to distinguish between what Google "designed" its recommendation algorithms to promote and what the algorithms "intentionally"[199] do promote. A machine learning algorithm, such as a deep neural network designed to teach itself to engage users,[200] could very well conclude based on historical and incoming data that ISIS videos increase user engagement, lead to more time on YouTube, encourage sharing of videos, etc., and therefore decide to promote ISIS videos over other, lawful videos. In fact, such an algorithm could conclude that radicalizing users was an effective way to achieve greater user engagement and ad distribution. That is, whatever the intent of Google engineers, the algorithm itself could intend to spread ISIS's message, in a way that is emphatically not content-neutral. Indeed, if the standard for "content-neutral" or "neutral intermediary" is only that all content is initially processed in the same way by the algorithm, then human beings are also content-neutral consumers and producers of content. Such a weak form of content neutrality is equivalent to proposing a literary critic is content-neutral because they read each new book, no matter the content, with the same eyes. Ultimately, a machine learning algorithm need not be content-neutral nor be a neutral intermediary, but rather it may demote or promote content as it independently decides how well the content meets the objectives of its programming, without any individual programmer intent to engage in unlawful conduct.

### *B. The* Gonzalez *Court's "Voluntary Inputs" Are Not Meaningfully Voluntary*

"Neutral tools," according to the courts, necessarily act on "voluntary inputs."[201] Voluntary inputs are "willingly provide[d]" by

196. *Gonzalez*, 2 F.4th at 895.

197. *Id.* (emphases added).

198. *Id.*

199. I am using the term loosely to mean the algorithm is promoting the content for a purpose (likely user engagement and ad views).

200. For example, one of the ostensible goals of user engagement is to generate ad revenue for YouTube's parent company Google.

201. *See Gonzalez*, 2 F.4th at 893 (quoting *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1172 (9th Cir. 2008)).

third parties to the neutral tool.[202] In the case of publishing tools or traditional surfacing tools like search engines,[203] those voluntary inputs are the content the user seeks to post or the query the user puts into the textbox.[204] However, in the case of surfacing tools like recommendation algorithms, it is not so obvious what the voluntary inputs are supposed to be. In response to this problem, the *Gonzalez* court made a subtle distinction, tweaking the wording of *Roommates.com*, to change a user's "voluntary inputs" to a user's "voluntary actions."[205]

The distinction between voluntary inputs and voluntary actions is necessary to finding that recommendations algorithms are neutral tools. The *Gonzalez* court equivocates on the word "voluntary," switching the meaning from inputs expressly made by users to any data the website can collect about the user.[206] YouTube's algorithms recommend videos "based on historical actions," the "content they have selected," and "other information about users."[207] This recommendation system, the court opined, is "certainly more sophisticated than a traditional search engine, which requires the users to type textual queries, [but] the core of the principle is the same: Google's algorithms select the particular content provided to a user based on that user's inputs."[208] Importantly dropping the "voluntary" from its construction, the court's apparent stance is that any information a website can gather about a user traversing the site is a voluntary input.[209]

Though the *Gonzalez* complaint simply alleges YouTube's recommendations are based on "what is known about the viewer," the reality is that what YouTube knows about a viewer may go far beyond what videos were watched or other inputs, direct or indirect, made to the site. A YouTube account is a Google account and, as mentioned above, Google can and does track users across the internet, on and off their own webpages.[210] Thus, what the court groups in "voluntary inputs" includes

---

202. *Id.* (quoting *Carafano v. Metrosplash.com*, 339 F.3d 1119, 1124 (9th Cir. 2003)).

203. It is not apparent that search engines, even when the internet was still a primordial soup, were ever so simple as the courts seem to think. *See* Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 COMPUT. NETWORKS & ISDN SYS. 107, 109–10 (1998) (Google founders Brin & Page, as Computer Science PhD students, describing their PageRank algorithm that powered the initial Google Search Engine).

204. *Gonzalez*, 2 F.4th at 893, 895.

205. *See id.* at 895; *Roommates.com*, 521 F.3d at 1172.

206. *See Gonzalez*, 2 F.4th at 893, 895.

207. *Id.* at 895.

208. *Id.* ("Google matches what it knows about users based on their historical actions and sends third-party content to users that Google *anticipates* they will prefer.") (emphasis added).

209. *See id.*

210. *See* discussion *supra* Section I.D.

both the express inputs of a user on a website and the incidental or indirect data accumulated about a user, either from using the website or possibly from other sites and third parties.

The approach to voluntary inputs used in *Gonzalez* is problematic because, as the Supreme Court recently held in *Carpenter v. United States*,[211] such data collection is not meaningfully voluntary.[212] In *Carpenter*, a Fourth Amendment search case involving "cell phone location information [that] is detailed, encyclopedic, and effortlessly compiled" automatically by cell phone companies, the Court modified its third-party voluntary disclosure doctrine.[213] The cell phone data at issue was collected "without any affirmative act on the part of the user beyond powering up" meanwhile carrying a cell phone was "indispensable to participation in modern society."[214] Thus, the Court concluded that the data collection was in "no meaningful sense . . . voluntar[y]" and therefore a search.[215]

Fourth Amendment jurisprudence does not govern what a private company can collect about an individual.[216] However, in the context of granting broad immunity to technology companies, the *Gonzalez* court never undertook an analysis of what makes inputs voluntary or willing, but rather assumed that if a website knows something about the user, it is voluntary even if the user is unaware the information is being collected.[217] The lack of analysis is important, because the core reasoning of *Roommates.com*, as stated by the court in *Gonzalez*, was that the website "*required* users to input discriminatory content as a prerequisite" to using the service.[218] A user's actions, and the knowledge of the user that algorithms obtain therefrom, are likewise required to be disclosed to YouTube as a condition of using the site.[219]

---

211. 138 S. Ct. 2206 (2018).

212. *Id.* at 2220.

213. *Id.* at 2216, 2220.

214. *Id.* at 2220.

215. *Id.* at 2220–21.

216. *United States v. Jacobsen*, 466 U.S. 109, 113 (1984).

217. *See Gonzalez v. Google LLC*, 2 F.4th 871, 895 (9th Cir. 2021), *cert. granted,* No. 21-1333, 2022 WL 4651229 (U.S. Oct. 3, 2022), *and cert. granted, sub nom. Twitter, Inc. v. Taamneh,* No. 21-1496, 2022 WL 4651263 (U.S. Oct. 3, 2022). For a detailed and compelling account of so-called "browsewrap," end-user agreements that allow web companies to impose one-sided contracts of adhesion on consumers, with the blessing of the courts, see Mark A. Lemley, *The Benefit of the Bargain*, 2023 WIS. L. REV. (forthcoming) (manuscript at 18–25), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4184946 [https://perma.cc/Z42G-5XZ4].

218. *Gonzalez*, 2 F.4th at 894 (emphasis added).

219. *Privacy & Terms*, GOOGLE, https://policies.google.com/privacy [https://perma.cc/9EWR-54DG] (last visited Oct. 30, 2022) ("We collect information about your activity in our services, which we use to do things like recommend a YouTube

Concomitant to the use of vast troves of unknowingly collected data about a user, is that recommendation algorithms are not surfacing content based only on what they know about the user but what they know about all users.[220] The use of multiple users' inputs compounds the problems of surfacing tools for neutral tool analysis, because instead of surfacing what it anticipates the user will want to watch, it may surface what people like the user have watched.[221] To justify their outcome, courts have settled on a definition of voluntary that results in a strange situation in which *nobody* is responsible for the information content of certain algorithmic recommendations. With multiple users' incidental data being aggregated by machine learning algorithms to generate recommendations, the reasoning articulated in *Roommates.com* of "neutral tools operating on voluntary inputs" becomes so attenuated as to lose all force.

### *C. Hypothetical: Applying* Gonzalez *'s "Neutral Tool" Analysis to* Roommates.com *'s Facts*

As a hypothetical, if we take the basic fact scenario from *Roommates.com*, adapting it to involve machine learning algorithms that operate on incidental user data, we come to the opposite conclusion on Section 230 immunity. Imagine a website (Flatmates.net), that matches people looking to let a room in their apartment with people looking to rent. Flatmates.net does not require users to expressly disclose their gender, sexuality, or familial status, however the website automatically collects data about how its users use the site. The data collected is quite extensive, including: which potential roommates they click, view, or hover over and for how long; what areas they are looking at to rent or let; which houses they look at; their usernames, registration emails, and (if they have bad security) passwords; IP addresses and locations; and any profile information such as pictures, listed interests or "looking for" blurbs, and descriptions of themselves. Flatmates.net has a tool, designed by perfectly angelic programmers to help facilitate the best possible match for each user, that recommends potential roommates. The recommendation tool uses deep learning algorithms to analyze a user's data—as well as the data collected in the same way from all other users—and suggest potential matches in a sidebar that is omnipresent while browsing the site.

---

video you might like. The activity information we collect may include: Terms you search for; Videos you watch; Views and interactions with content and ads; Voice and audio information when you use audio features; Purchase activity; People with whom you communicate or share content; Activity on third-party sites and apps that use our services; Chrome browsing history you've synced with your Google Account.").

220. *See* discussion *supra* Section I.C.

221. *See* discussion *supra* Section I.C.

The algorithm does its job by categorizing users in unknowable ways and by imperceptible inferences based on historical and continuously incoming data. With such black box categorizations, it starts matching people. One user, say, Karen—a straight, single woman—notices after a while that all the roommates Flatmates.net recommends to her in the sidebar are all women, all straight, and all single. Thus, the recommendation algorithm has, we can surmise, begun to "steer users based on discriminatory criteria."[222]

Under the court's *Gonzalez* reasoning, Flatmates.net would be immune from suits arising from the illegal discriminatory steering. After all, Flatmates.net's recommendation algorithm was a neutral tool—Flatmates.net simply "provided a neutral platform that did not specify or prompt the type of content to be submitted, nor determine particular types of content its algorithms would promote."[223] Despite the fact that Karen never expressly searched based on discriminatory criteria, her "voluntary actions" guided the algorithm into unlawfully steering her on a discriminatory basis. Without intent and the attendant express requirement of disclosure, Flatmates.net is not the "information content provider" of its own recommendations.

Here, the "neutral tool" operating on "voluntary inputs" has engaged in the exact unlawful conduct at issue in *Roommates.com*, only this time Flatmates.net is immune under the court's overly broad reading of Section 230. The sole differences between the two scenarios are that there can be no allegation of intent on the part of the developers and that the required disclosure of characteristics was implicit rather than express. Despite courts' apparent beliefs to the contrary, machine learning algorithms are capable of identifying discrimination as an optimal way to meet their goal and then engaging in it. Clearly, allowing machines to discriminate—or otherwise behave unlawfully[224]—in place of humans was not the intent of Section 230.

## III. UNDER A PROPOSED DISCRETIONARY TEST, MACHINE LEARNING ALGORITHMS CAN BE INFORMATION CONTENT PROVIDERS

Certain algorithmic recommendation systems are, and ought to be treated as, information content providers for the purposes of Section 230. This Part first proposes the discretionary test as an alternative framework to neutral tool analysis that better fits the current challenges presented by surfacing algorithms. Next, the discretionary test is used to analyze cases

222. *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167 (9th Cir. 2008).

223. *Gonzalez*, 2 F.4th at 895.

224. *E.g.*, by recommending terrorist propaganda videos to those vulnerable to radicalization or heroin groups to recovering addicts.

decided under neutral tool analysis to show how it arrives at a more intuitive and statutorily faithful result. While the neutral tool framework is capable of being applied reasonably to publishing tools, it is inadequate to help in the analysis of surfacing tools, especially those that do not require express user input, like some algorithmic recommendation systems. Instead of (mis)adapting a test developed when the capabilities of internet service providers were much less sophisticated, courts should acknowledge that the original construction given to Section 230 has proved unworkable in the face of increasingly complex and advanced technology.

To better analyze algorithmic information content creation under Section 230, it is helpful to consider what constitutes information creation generally. Most people would understand a person recommending a book to a friend to be providing information other than what is contained in the book. Similarly, we understand there is a difference between a librarian personally recommending a book and a librarian telling us where a book is located in the library. The location of the book in the library is information that already exists, but the personalized recommendation, based on information the librarian gleans about the reader and knows about the catalogue, is new information synthesized from old. Indeed, asking a different librarian is likely to produce a different recommendation based on the same information. The entire tradition of science is premised on this idea that unique synthesis of many disparate pieces of information creates new information.[225] This understanding of information creation is more faithful to the text of the statute, which merely provides that an "information content provider" is "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."[226] Nothing in the text renders the definition inapplicable to the use of algorithms, which are otherwise responsible for the creation or development of information, merely because they are algorithms.

With that understanding of information content creation, the test for whether an algorithm is creating or developing information content becomes simpler: if it is possible to take in the same information and return different outputs based on a choice of values, judgments, or inferences—*i.e.*, if the output is discretionary—then the output is new

225. *See, e.g.*, KEITH E. STANOVICH, HOW TO THINK STRAIGHT ABOUT PSYCHOLOGY 142 (10th ed. 2013); Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339, 342 (2017) ("Data become valuable through analysis, turning unstructured bits and bytes into information and derived information—*i.e.*, applying reasoning mechanisms to create new information that cannot be gathered directly from the data—in order to turn them into actionable information, both descriptive as well as predictive.") (italics added).

226. 47 U.S.C. § 230(f)(3).

information content. Whether the values, judgments, or inferences are those of a human or those of an algorithm is irrelevant to the inquiry. A librarian providing the location of a book is not creating information content, because there is no room for discretion. In contrast, a librarian providing a book recommendation is creating information content, because it depends on how they weigh values and make inferences from the information they know about a patron combined with information they have about the books in the library's catalogue. Importantly, discretion in *what* information is *returned*, not discretion in *how* information is *displayed*, is the proper test for the creation or development of information.

Likewise, an algorithm creates new information when the content produced could have been different with different values, judgments, or inferences. For instance, Google Maps can drop a pin on any address in the world. There is no room for discretion, because there is only one possible output for the input.[227] In contrast, when one requests not just a map location, but directions from one location to another, Google Maps is creating or developing information. There are an infinite number of ways to get from one location to another (as the crow flies—only slightly less than that when limited to roads), yet Google Maps picks and displays one particular route, based on its understanding of mathematical models, traffic conditions collected from millions of unwitting smart phone users, and other considerations.[228] What Google Maps—or rather its machine learning algorithm—is asserting is that it believes this to be the best route for the user to take. However, it is neither the only possible route nor necessarily the objectively "best" one.

As an example, if we again take up the facts of the Flatmates.net hypothetical[229] and apply this new discretionary test to them, we get a proper result in line with the original decision in *Roommates.com*. The machine learning algorithm is taking the mountainous information it collects about Karen and—based on the values, judgments, and inferences it has developed through analyzing the data amassed from a myriad of individuals—makes specific roommate recommendations to Karen. Though it is unclear for certain why the algorithm is recommending only single, straight, women (one reason why a narrower reading will not automatically impose liability on those who deploy machine learning algorithms),[230] there is an obvious deduction that some of the judgments

227. Or if there is more than one location corresponding to a particular address, Google Maps will ask the user to resolve the dilemma.

228. Johann Lau, *Google Maps 101: How AI Helps Predict Traffic and Determine Routes*, GOOGLE (Sept. 3, 2020), https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes [https://perma.cc/X88F-E9MV].

229. *See* discussion *supra* Section II.C.

230. *See* discussion *infra* Part IV.

and inferences the algorithm made were based on impermissible criteria under the Fair Housing Act. For the purposes of the discretionary test, however, what matters is that the algorithm could have recommended any number of roommates but instead exercised its discretion to choose those particular roommates based on information it knew about both Karen and the potential roommates, thereby creating or developing new information.

Similarly, in the case of *Dyroff*, the machine learning algorithm's decision to recommend heroin groups to what it had inferred was a recovering drug addict—"based on the content of their posts and other attributes"[231]—was a discretionary call based on what it had determined would increase user satisfaction or engagement with the website.[232] YouTube's recommendation algorithm in *Gonzalez* fares no better when it used the voluminous, involuntary inputs from its users in deciding to surface videos supporting ISIS by "anticipating" users' preferences.[233] Again, whether the recommendations alone are enough to impose liability on Google is a separate question from who created or developed the recommendations as information content.

In contrast to those cases, take *Marshall's Locksmith Service v. Google*,[234] in which the Court of Appeals for the District of Columbia Circuit affirmed the dismissal of a market conspiracy suit on the basis of Section 230 immunity.[235] In that case, plaintiff locksmiths alleged that Google scraped location information from scam locksmiths' fake webpages and then displayed the scam locksmiths as pinpoints on the map as local businesses, thereby harming legitimate, local locksmiths' businesses.[236] For the purposes of Section 230 immunity, two specific allegations are relevant. First, plaintiffs alleged that if a fake website contained a fake address, Google would scrape it and display it on a map as a pinpoint, thereby creating or developing the address information.[237] Second, plaintiffs alleged that if a fake website contained incomplete location information (such as a city or area code), Google would

231. *Dyroff v. Ultimate Software Grp., Inc.*, 934 F.3d 1093, 1095 (9th Cir. 2019).

232. *See id.*

233. *Gonzalez v. Google LLC*, 2 F.4th 871, 895 (9th Cir. 2021), *cert. granted,* No. 21-1333, 2022 WL 4651229 (U.S. Oct. 3, 2022), *and cert. granted, sub nom. Twitter, Inc. v. Taamneh,* No. 21-1496, 2022 WL 4651263 (U.S. Oct. 3, 2022).

234. *Marshall's Locksmith Serv. Inc. v. Google, LLC*, 925 F.3d 1263 (D.C. Cir. 2019).

235. *Id.* at 1265.

236. *Id.* at 1265–66.

237. *Id.* at 1269.

randomly choose[238] an address within the relevant area to display a map pinpoint of the supposed locksmith.[239]

The court affirmed Section 230 immunity on both claims, using neutral tool analysis.[240] In the first instance, the court correctly reasoned that a choice in how to display information provided[241] by another information content provider did not create or develop information.[242] However, in the second instance, the court employed neutral tool analysis to find Section 230 immunity, stating "the location of the map pinpoint is *derived* from the scam-locksmith information: its location is *constrained* by the underlying third-party information."[243]

The discretionary test articulated in this Comment reaches the same conclusion on the first claim. Because there was no exercise of discretion dependent on particular values, judgments, or inferences in *what* address was to be displayed, only discretion in *how* to display the provided address, there was no creation or development of information. However, the discretionary test comes to a different result for the second claim. While the exact address of the pinpoint on the map was "constrained by the underlying third-party information," Google still exercised discretion in where exactly within the relevant area to place the marker. Google's choice whether to place the pinpoint in the geographic center, the population center, completely randomly, or at location close to the current user, was a choice made based on what the company sought to achieve and a choice that provided independent information (though false) to users of the site. Thus, while the discretionary test casts a broader net than neutral tools analysis for what constitutes an "information content provider" under the statute, it hews closer to a textual understanding of the terms as well as the original congressional intent to promote voluntary censorship of obscene content on the web. In doing so, it restricts the broad grant of Section 230 immunity courts have been handing out to companies that deploy machine learning and other algorithms.

238. In a forfeited argument, relevant to the current discussion, plaintiffs claimed that Google did not merely select these addresses randomly, but rather chose a location close to the user who was currently searching. *Id.* at 1271.

239. *Id.* at 1270.

240. *Id.* at 1270–72.

241. *See* discussion *supra* note 160.

242. *Marshall's Locksmith Serv.*, 925 F.3d at 1269.

243. *Id.* at 1270 (emphases added).

## IV. LIABILITY FOR MACHINE LEARNING ALGORITHMS ABSENT SECTION 230 IMMUNITY

The largest policy reason against a narrower construction of Section 230, highlighted both in the legislative history of the CDA and by the courts, is that too broad of liability for internet companies will inhibit what has become the largest repository of free expression and information in the history of the world.[244] However, internet service providers like Google reap enormous benefits from their substantial predominance and control over large swaths of the internet.[245] It is exceedingly unlikely that reducing the barriers to recovery for those injured by Google's highly lucrative algorithms will cause Google to decide to leave the market, or even significantly change its business. Hopefully, narrowing the scope of immunity will incentivize such companies to reduce the potential harm their platforms can inflict and make whole those who have been injured. Further, returning Section 230 to its proper scope is unlikely to precipitate a tidal wave of liability for tech companies, because the traditional barriers to liability when dealing with machine learning algorithms are already high.

For instance, while intent or knowledge should not be read into the definition of "information content provider," the issue will often arise in many causes of action. For example, the plaintiffs in *Gonzalez* had to allege that Google acted knowingly in recommending the ISIS propaganda videos. Even without Section 230 immunity, proving knowledge to a sufficient degree through the intermediary of a machine learning algorithm may prove very challenging, if not insurmountable, for plaintiffs. Thus, knowledge and intent requirements may still stymy efforts to hold Google and other internet service providers liable for their surfacing algorithms. An analysis of machine learning liability in the United Kingdom, beyond the remit of Section 230, similarly concluded that "the current legal basis [for liability] would be faced with problems of shared responsibility and lack of knowledge about how the machine learning was trained and the basis on which it makes its decisions."[246]

---

244. Keenan Mayo & Peter Newcomb, *How the Web Was Won*, VANITY FAIR (Jan. 7, 2009), https://www.vanityfair.com/news/2008/07/internet200807 [https://perma.cc/EGL5-BGXC].

245. *See* Graham & Elias, *supra* note 117; *see also* Rothchild, *supra* note 110, at 568–69.

246. Chris Reed, Elizabeth Kennedy & Sara Nogueira Silva, *Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning* 26 (Queen Mary Univ. of London, School of Law Legal Studies Research Paper, Paper No. 243/2016) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2853462 [https://perma.cc/S2B2-QAYV] ("Resolving these under the current law would be likely to require imputations of knowledge to the persons using or owning the technology which are obviously untrue.").

Causation will also be a factor in some claims that may frustrate plaintiffs' attempts to hold algorithms' creators responsible for torts and other alleged misconduct.[247] In the *Gonzalez* case, the chain of causation between deploying a YouTube video recommendation algorithm and radicalizing a man in Belgium was sufficiently attenuated and unforeseeable that it may not have been within the scope of the risk assumed. Finally, there may also be a freedom of speech issue regarding recommendation algorithms "speaking" for their creators.[248] Because of the difficulties machine learning algorithms present for many normal requirements of liability, a narrower, more textually faithful reading of Section 230 would not unduly limit the freedom of the internet.

## CONCLUSION

The cramped construction of "information content provider" under neutral tool analysis has improperly excluded machine learning algorithms from the realm of Section 230 information content providers, granting companies broader immunity than the text, purpose, or congressional intent of the CDA warrant. Because of the unworkability of neutral tool analysis in the face of rapidly developing artificial intelligence technology, a more robust test is overdue. The discretionary test is a more appropriate standard which finds information to have been created or developed when a person or machine exercises independent values, judgments, or inferences to produce an output not necessitated by the inputs. Though a more expansive reading of "information content provider" under Section 230 would result in less immunity for internet companies, the independent barriers to litigation for machine learning algorithms would still afford such companies significant protection, while allowing litigants the opportunity for discovery and fair recovery. Courts

247. *Id.* at 13 ("[W]here the machine learning technology's decision-making element comprises a neural network, or some similar technology, it will be difficult and perhaps impossible to explain how the technology came to its decision, and thus how the loss or damage was caused.").

248. I think that is a misguided view of deep learning algorithms, as the so-called "speech" is completely disconnected from the knowledge, direction, or understanding of the algorithm's creators. That position is akin to attributing the speech of a child to the parent because the parent intended to create a child that could speak. *But see* Ellen Smith Yost, *Social Support for Terrorists: Facebook's "Friend Suggestion" Algorithm, Section 230 Immunity, Material Support for Terrorists, and the First Amendment,* 37 SANTA CLARA HIGH TECH. L.J. 301, 328 (2021). However, if the recommendations of machine learning algorithms are protected speech, as Google suggests, *cf.* Eugene Volokh & Donald M. Falk, *Google: First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL'Y 883 (2012), it leads to a telling conflict between asserting immunity under Section 230—because there was no creation or development of information—and simultaneously claiming that the same information is Google's own protected speech.

should eschew neutral tool analysis in favor of the more workable discretionary test for determining when an internet service provider deploying advanced algorithms is also an "information content provider."

* * *