Note

Evolving Online Terrain in an Inert Legal Landscape: How Algorithms and AI Necessitate an Amendment of Section 230 of the Communications Decency Act

Ellison Snider[*]

INTRODUCTION

On January 6, 2021, an irate crowd stormed the nation's Capitol, attempting to "stop the steal"[1] and block the confirmation of then-President-elect Joe Biden. Threats were made to elected officials,[2] historical government property was damaged,[3] and seven people died as a result.[4] Shock rippled through a grieving country, leaving many to wonder how this type of politically frustrated violence could happen in a country so distinctly characterized by its democratic identity. The last time the Capitol

* Managing Editor, *Minnesota Law Review*, Volume 107; J.D. Candidate 2023, University of Minnesota Law School. I would like to thank Professor Alan Z. Rozenshtein for his expert guidance and feedback on this Note; Marra Clay, Calvin Lee, and the *MLR* Note and Comment Department for their efforts and encouragement; and Dylan Saul and the *MLR* Managing Department for their exacting improvements to this Note. My deepest thanks, also, to Kelsee and my family for reminding me who I am and rooting for me always. Copyright © 2023 by Ellison Snider.

1. *See, e.g.*, Sheera Frenkel, *The Rise and Fall of the 'Stop the Steal' Facebook Group*, N.Y. TIMES (Nov. 5, 2020), https://www.nytimes.com/2020/11/05/technology/stop-the-steal-facebook-group.html [https://perma.cc/3E4Y-SWTN] (explaining the origins of the "stop the steal" slogan).

2. *See, e.g.*, Ben Mathis-Lilly, *Trump Says Rioters Wanting to Kill Mike Pence on Jan. 6 Was "Common Sense,"* SLATE (Nov. 12, 2021), https://slate.com/news-and-politics/2021/11/trump-common-sense-jan-6-kill-mike-pence.html [https://perma.cc/H5WF-N8ZY].

3. *One Year Since the Jan. 6 Attack on the Capitol*, U.S. DEP'T OF JUST. (Dec. 30, 2021), https://www.justice.gov/usao-dc/one-year-jan-6-attack-capitol [https://perma.cc/3DKD-9TUU] (detailing the damage done to the Capitol and its surrounding grounds).

4. *See* Chris Cameron, *These Are the People Who Died in Connection with the Capitol Riot*, N.Y. TIMES (Jan. 5, 2022), https://www.nytimes.com/2022/01/05/us/politics/jan-6-capitol-deaths.html [https://perma.cc/CKB8-5F5M].

faced such an attack was when British troops sieged and burned the building in 1814.[5] At the time, the British weapons of choice were simple rockets that were impossible to accurately aim.[6] On January 6th, the insurrectionists wielded a much more sophisticated and detrimental technology—the internet.[7]

Though much of the country shuddered in disbelief,[8] there was a small subset of internet users who saw the writing on the wall. Users of the platform Parler knew the January 6th insurrection was imminent.[9] Parler's minimal moderation of user-posted content distinguishes it from competing platforms.[10] A

5. *See generally* Joel Achenbach, *In 1814, British Forces Burned the U.S. Capitol*, WASH. POST (Jan. 6, 2021), https://www.washingtonpost.com/history/2021/01/06/british-burned-capitol-1814 [https://perma.cc/E48C-RAXQ].

6. *Id.*

7. *See, e.g.*, Sheera Frenkel, *The Storming of Capitol Hill Was Organized on Social Media*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html [https://perma.cc/6CFV-XDGC]; Rebecca Heilweil & Shirin Ghaffary, *How Trump's Internet Built and Broadcast the Capitol Insurrection*, VOX (Jan. 8, 2021), https://www.vox.com/recode/22221285/trump-online-capitol-riot-far-right-parler-twitter-facebook [https://perma.cc/2XKU-7C5Y].

8. *See* John Gramlich, *A Look Back at Americans' Reactions to the Jan. 6 Riot at the U.S. Capitol*, PEW RSCH. CTR. (Jan. 4, 2022), https://www.pewresearch.org/fact-tank/2022/01/04/a-look-back-at-americans-reactions-to-the-jan-6-riot-at-the-u-s-capitol [https://perma.cc/5K7D-2MLL] (reporting that forty-nine percent of American survey participants disclosed feeling either anguished, horrified, or shocked or, alternatively, surprised and concerned for the country when they heard about the January 6, 2021, insurrection).

9. *See* Sway, *If You Were on Parler, You Saw the Mob Coming*, N.Y. TIMES (Jan. 7, 2021), https://www.nytimes.com/2021/01/07/opinion/sway-kara-swisher-john-matze.html [https://perma.cc/8USQ-SNTF].

10. *Compare Values*, PARLER, https://parler.com/values.html [https://perma.cc/7VRT-FM6U] ("Our objective Community Guidelines focus on Supreme Court and [Federal Communications Commission] precedence."), *with Facebook Community Standards*, META, https://transparency.fb.com/policies/community-standards [https://perma.cc/AF8R-CLY3] ("These standards are based on feedback from people and the advice of experts in fields like technology, public safety, and human rights.").

self-described "neutral town square,"[11] Parler serves as the platform[12] of choice for the alt-right.[13] Before January 6th, Parler boasted an especially hands-off approach to content moderation on its site.[14] After the insurrection, however, this laissez-faire approach led to Parler's "deplatforming."[15] After fortifying its content moderation practices[16] to Big Tech's[17] liking, Parler is back online.[18]

---

11. Sway, *supra* note 9 (quoting Parler's former CEO, John Matze).

12. This Note uses the word "platform" to broadly encompass consumer-facing websites like Google, Facebook, and YouTube that primarily host user-developed content. These are privately owned companies, and some of them, such as Google, provide services and products besides the hosting of user-developed content that is the subject of this Note.

13. Amanda Silberling, *Right-Wing Social App Parler Raises $20M in Funding*, TECHCRUNCH (Jan. 7, 2022), https://techcrunch.com/2022/01/07/right-wing-social-app-parler-raises-20m-in-funding [https://perma.cc/R27L-49KK] (describing Parler as a "conservative social media platform" that prides itself on being against "Big Tech, Big Government, and cancel culture"). Alt-right conservatives are drawn to Parler because, often, this political constituency is extremely concerned with the Communications Decency Act of 1996 (CDA 230) and censorship. *See, e.g.*, Meysam Alizadeh, Fabrizio Gilardi, Emma Hoes, K. Jonathan Kluser, Mael Kubli & Nahema Marchal, *Content Moderation as a Political Issue: The Twitter Discourse Around Trump's Ban*, 2 J. QUANTITATIVE DESCRIPTION: DIGIT. MEDIA 1, 18 fig.6(a) (2022) (showing that conservative Twitter users are far more likely to post about censorship and CDA 230 than their left-leaning counterparts).

14. If content was in violation of the law or the platform's terms, the content was sent to a jury of the creator's peers—five Parler users—and a four out of five majority was needed to remove content. Sway, *supra* note 9. All decisions were made by humans. *Id.*

15. Apple removed Parler from its app store, Google deleted Parler from Google Play, and Amazon dropped Parler from its web hosting service. Sway, *One Year After the Jan. 6 Attack, Parler's C.E.O. Grapples with Big Tech and Trump*, N.Y. TIMES (Jan. 6, 2022), https://www.nytimes.com/2022/01/06/opinion/sway-kara-swisher-george-farmer.html [https://perma.cc/X4XB-5CNS].

16. Notably, Parler now heavily relies on AI to moderate content on its site. *Id.*

17. For purposes of this Note, Big Tech is defined as the five largest internet companies—Google, Amazon, Meta, Apple, and Microsoft. J. Clement, *Google, Amazon, Meta, Apple, and Microsoft (GAMAM)—Statistics & Facts*, STATISTA (Oct. 18, 2022), https://www.statista.com/topics/4213/google-apple-facebook-amazon-and-microsoft-gafam/#dossierKeyfigures [https://perma.cc/7FPX-3YZ9].

18. According to Parler CEO George Farmer, Apple wanted two filters placed on the Parler app: a "not safe for work" filter for explicit content, and a "troll filter" for racist and abusive content. Sway, *supra* note 15. As a result, content viewed on the app is shielded with a warning, but Parler users are still able to access the content without the warning on a web browser. *Id.*

Parler's laissez-faire content moderation approach and consequently poor reputation made the platform an obviously accountable culprit of the insurrection. But mainstream platforms like Facebook and Twitter were not absolved of their responsibility.[19] These platforms are a central feature of the modern internet and platforms' content moderation practices, including their reliance on algorithms and their employment of artificial intelligence (AI),[20] contribute to what makes the internet so complicated to regulate.[21]

A powerful piece of legislation—a safe harbor for the internet—was enacted to support online innovation before algorithms and AI began shaping the internet. Section 230 of the Communications Decency Act of 1996 (CDA 230) has been credited with "creating the [i]nternet."[22] CDA 230 immunizes platforms like Parler from liability for user-developed content on their sites.[23] Modern judicial expansion of the CDA 230 shield has distorted the law[24] and enabled on the one hand, "a form of social life in which unjust barriers of rank and privilege are dissolved, and in

---

19. *See* Craig Silverman, Craig Timberg, Jeff Kao & Jeremy B. Merrill, *Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading up to Jan. 6 Attack, Records Show*, PROPUBLICA (Jan. 4, 2022), https://www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show [https://perma.cc/BD5H-5RGC]; Graeme Massie, *A Timeline to Insurrection: The Trump Tweets That Security Experts Say Led to the Capitol Riots*, INDEPENDENT (Jan. 18, 2021), https://www.independent.co.uk/news/world/americas/us-election-2020/trump-tweets-attacks-capitol-violence-b1786246.html [https://perma.cc/CNX6-CCGD].

20. Artificial intelligence, commonly known as "AI," is a broad concept to describe compositions of algorithms that can self-modify and create new algorithms "in response to learned inputs and data as opposed to relying solely on the inputs it was designed to recognize as triggers." Kaya Ismail, *AI vs. Algorithms: What's the Difference?*, CMS WIRE (Oct. 26, 2018), https://www.cmswire.com/information-management/ai-vs-algorithms-whats-the-difference [https://perma.cc/6QHM-QUNA].

21. *See* Dipayan Ghosh, *Are We Entering a New Era of Social Media Regulation?*, HARV. BUS. REV. (Jan. 14, 2021), https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation [https://perma.cc/JPZ5-CT45] (discussing the "key reasons these issues are so difficult to untangle").

22. *See* JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET (2019) (examining how Congress passed CDA 230, how it shaped the internet, and the costs and benefits of liability immunity for sites hosting user-developed content).

23. *See generally* 47 U.S.C. § 230(c).

24. *See infra* Part II.B.

which ordinary people gain a greater say,"[25] and on the other, "a toxic swamp of hate, conspiracy threats, and anger."[26] The internet of today is unrecognizable from the internet of 1996 and, importantly, the mechanics of the internet and reach of online speech are unfathomable to the greatest innovators and lawmakers of 1996.[27] The introduction of platforms and algorithms that amplify ordinary speakers and moderate user-developed content has spurred labyrinthine debate, particularly concerning the expansive immunity afforded to platforms under CDA 230[28] and the need for regulation and legal solutions to the uniquely puzzling issue of online speech regulation.

It is imperative to determine how to regulate online speech because of its consequences in the offline world, such as harming platform users' mental health and the erosion of democracy.[29] Just as platforms and algorithms have facilitated the internet's metamorphosis since CDA 230's enactment, they can be expected to continue shaping the online, and thus, the offline world.[30] Developing a regulatory regime for online speech requires a nuanced balancing of interests, namely, respect for freedom of expression, empowerment of technological innovation, and protection of individual and societal safety and wellbeing.

25. Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 35 (2004).

26. Sway, *supra* note 9.

27. *See* Mary Graw Leary, *The Indecency and Injustice of Section 230 of the Communications Decency Act*, 41 HARV. J.L. & PUB. POL'Y 553, 557 (2018) (discussing the unintended regime of "absolute immunity" for defendant websites).

28. *See, e.g.*, Morgan Weiland, *CDA 230 Problems: Do Algorithms Threaten to Undermine Speech Protections*, CTR. FOR INTERNET & SOC'Y (June 22, 2016), http://cyberlaw.stanford.edu/blog/2016/06/cda-230-problems-do-algorithms-threaten-undermine-speech-protections [https://perma.cc/K7WH-6G3H] (describing arguments developed through an email exchange between Weiland and Professor Eugene Volokh about the functional transformation of internet intermediaries and the impact on First Amendment and CDA 230 protections).

29. *See infra* Part I.A.3.

30. *See, e.g.*, Sway, *The Metaverse: Expectations vs. Reality*, N.Y. TIMES (Nov. 11, 2021), https://www.nytimes.com/2021/11/11/opinion/sway-kara-swisher-jaron-lanier.html [https://perma.cc/GX4P-QX6K] (interviewing Jaron Lanier, an early pioneer of virtual reality, about Meta's quest to build the metaverse and how these "technologies will continue to shape our lives").

This is not a modest task, but there are already several significant efforts to balance these interests and restrain platforms from absolute immunity under CDA 230.[31]

There is a vacuum in the legal landscape to regulate online speech of the modern internet. Part I of this Note describes key characteristics of the modern internet that impede the effectuation of traditional speech regulation frameworks and exacerbate offline impact. This Part then describes the law of the internet: CDA 230. Part II explores one scholar's theory of a new modern speech regulation framework succeeding the internet's rise. Part III argues that understanding this modern speech regulation framework exemplifies CDA 230's incompatibility with the modern internet and analyzes judicial decisions that exhibit the judiciary's difficulty addressing algorithmic harm under CDA 230. Part IV then synthesizes and critiques various scholarly perspectives and proposed solutions to better enable online speech regulation through slight but impactful alterations to CDA 230's statutory text or by introducing a new reasonableness standard for online algorithmic action. This Note concludes with the resignation that, though these proposals are imperfect, their key features, together, reveal a path towards the regulation of online speech, and the platforms that host and propel it.

I. THE EVOLVING ONLINE TERRAIN AND AN INERT LEGAL LANDSCAPE

This Part seeks to explain the complicated nature of regulating the internet by juxtaposing the intricate mechanics and remarkable offline impact of the modern internet with its comparatively archaic and naïve regulatory context. Given the internet's rudimentary regulation, it is, predictably, difficult to contain Big Tech's power and influence. This Part first describes the modern internet and explains the role of algorithms as critical features of the platforms that host considerable online speech. Next, this Part briefly explains CDA 230 as it was enacted by

31. *See, e.g.*, Meghan Anand, Kiran Jeevanjee, Daniel Johnson, Quinta Jurecic, Brian Lim, Irene Ly, Matt Perault, Etta Reed, Jenna Ruddock, Tim Schmeling, Niharika Vattikonda, Brady Worthington, Noelle Wilson & Joyce Zhou, *All the Ways Congress Wants to Change Section 230*, SLATE (Mar. 23, 2021), https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html [https://perma.cc/AW6H-9QTH] (tracking all congressional attempts to amend CDA 230).

the legislature and as it has been interpreted by the courts, noting that, as the interpretation of CDA 230 immunity has broadened, so has the reach and impact of online speech.

A. THE MODERN INTERNET AND THE AGE OF ALGORITHMS

We live in an Algorithmic Society—"a society organized around social and economic decision-making by algorithms, robots, and AI . . . ."[32] Dominant in the online landscape are popular, privately-owned platforms.[33] These platforms are the prevailing actors in the internet's digital infrastructure and serve as intermediaries between traditional governments and ordinary people.[34] They are fueled by Big Data,[35] built by algorithms, and run by AI.[36] True appreciation for the challenges of regulating online speech requires an understanding of the modern internet and these platforms' complexity.

1. AI and Platform Functionality

In 1996, only 20 million Americans used the internet,[37] and most user-developed content was shared on simple online forums

32. Jack M. Balkin, Lecture, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1219 (2017).

33. *See* Jack M. Balkin, Essay, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1182–83 (2018) (discussing the development of private governance by companies like Twitter and YouTube).

34. *Id.*

35. *Id.* at 1154–57 ("Big Data is crucial to the use and development of algorithms and [AI]. Algorithms and AI are the machines; Big Data is the fuel that makes the machines run."). For purposes of this Note, Big Data includes "Information assets characterized by such a High Volume, Velocity, and Variety to require specific Technology and Analytical Methods for its transformation into Value." Andrea De Mauro, Marco Greco & Michele Grimaldi, *What Is Big Data? A Consensual Definition and a Review of Key Research Topics*, 1644 AIP CONF. PROC. 97, 103 (2015). That is, Big Data describes colossal swaths of data so quickly, massively, and diversely generated or collected so as to only be valuable with the development of sophisticated technology and interpretative methods to comprehend its message and impact.

36. Balkin, *supra* note 33, at 1151 (pointing to the use of algorithms and AI to govern populations).

37. Farhad Manjoo, *Jurassic Web*, SLATE (Feb. 24, 2009), https://slate.com/technology/2009/02/the-unrecognizable-internet-of-1996.html#:~:text=In%201996%2C%20just%2020%20million,the%20new%20medium%20was%20for [https://perma.cc/2GA5-SHA3] (comparing that number to subscribers of satellite radio in 2009).

and digital bulletin boards.[38] Today, platforms like Facebook, TikTok, and YouTube facilitate engagement among users and rapidly spread user-developed content. Platforms have transformed the marketplace of ideas by providing ordinary people with access to a large audience.[39] For example, Facebook hosts 2.9 billion monthly active users,[40] TikTok has 1 billion monthly users,[41] and YouTube has 2.1 billion users worldwide,[42] averaging 32.8 million subscribers.[43] Technology has cheaply and efficiently facilitated information sharing and, as a result, the public is inundated with user-developed content.[44] Algorithms and AI were developed to make something usable out of this information by moderating user-developed content and curating platform experience based on user interest.[45]

At their simplest, algorithms are sequenced steps a computer takes to analyze a data input and create a data output.[46] An algorithm's rules can be straightforward to sort and search information in a structured, automatic process when it encounters a trigger, such as an "if, then" sequence.[47] For example, a

---

38. *Id.* (reflecting on the beginning of "blogging" and simple, reverse-chronological web pages).

39. *See* Sofia Grafanaki, *Platforms, the First Amendment and Online Speech: Regulating the Filters*, 39 PACE L. REV. 111, 124 (2018) (referencing how technology has made it "cheap, quick, and easy" to make content available online).

40. *See* S. Dixon, *Countries with the Most Facebook Users 2022*, STATISTA (July 26, 2022), https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/#:~:text=With%20around%202.9%20billion%20monthly,most%20popular%20social%20media%20worldwide [https://perma.cc/73CU-T8C7].

41. *See* Daniel Ruby, *36 TikTok Statistics 2023: How Many Users Are There!*, DEMANDSAGE (Jan. 20, 2023), https://www.demandsage.com/tiktok-user-statistics [https://perma.cc/8URS-ED4A].

42. *See* L. Ceci, *YouTube—Statistics & Facts*, STATISTA (Nov. 23, 2022), https://www.statista.com/topics/2019/youtube/#topicHeader__wrapper [https://perma.cc/MRW5-NSKE].

43. *YouTube User Analytics/Statistics for YouTube*, SOC. BLADE, (Feb.–Mar. 2022), https://socialblade.com/youtube/user/youtube/monthly [https://perma.cc/PEL2-RWMA].

44. Grafanaki, *supra* note 39, at 125 (explaining that algorithms help audiences navigate content that is otherwise impossible to sort through).

45. *Id.*

46. *See* THOMAS H. CORMEN, CHARLES E. LEISERSON, RONALD L. RIVEST & CLIFFORD STEIN, INTRODUCTION TO ALGORITHMS 5 (3d ed. 2009).

47. *See* Philip Sales, *Algorithms, Artificial Intelligence, and the Law*, 105 JUDICATURE 22, 24 (2021) (distinguishing algorithm analysis and artificial intelligence).

simple algorithm may be used to sort a list of film titles in alphabetical order.

Algorithms can be much more complicated, however. AI is a broad concept, describing compilations of algorithms that self-modify and generate new algorithms in response to learned inputs and data.[48] One common form of AI is machine learning. Machine learning technology draws inferences based on user data and generates predictions; outputs are made based on the technology's own rules, derived from generalizations the computer has drawn from input data.[49] Machine learning pulls from generalized examples to come up with its own rules for output.[50] These outputs are called predictions[51] and these predictions shape much of an internet user's information diet.[52] For example, YouTube's recommendation engine employs AI to note each user's viewing habits and gauge their interest based on signals such as clicks, watch time, survey responses, sharing, likes, and dislikes.[53] YouTube's engine then sorts through billions of pieces of video content to recommend similar content or to recommend content that users' with similar viewing habits engaged with.[54]

When platforms emerged onto the internet, they perceived themselves, or at least held themselves out to the public, as

---

48. *See* Ismail, *supra* note 20.

49. *See, e.g.*, Gabriel Nicholas, *Explaining Algorithmic Decisions*, 4 GEO. L. TECH. REV. 711, 714 (2020) (discussing different "well-defined computational procedures").

50. *Id.*

51. *Id.*

52. *See* Marielle DeVos, *The Echo Chamber Effect: Social Media's Role in Political Bias*, INST. FOR YOUTH IN POL'Y (June 21, 2021), https://www.yipinstitute.com/article/the-echo-chamber-effect-social-medias-role-in-political-bias [https://perma.cc/48Y8-RPWP] ("Search engine optimization and different algorithms limit 'selection processes' because they constantly suggest content that is similar to what the user has already shown preference for. Platforms suggest groups and content to users based on observed preferences, so if a user shows preferences for conservative news sources, algorithms will suggest more conservative accounts and groups. . . . If a user shows preference for liberal news sources or content, the range of content they are shown will shift farther to the left as the algorithm observes their preferences.") (citations omitted).

53. Cristos Goodrow, *On YouTube's Recommendation System*, YOUTUBE (Sept. 15, 2021), https://blog.youtube/inside-youtube/on-youtubes-recommendation-system [https://perma.cc/G2CM-X3ZF] (discussing how YouTube's recommendation system operates).

54. *Id.*

merely tech companies.[55] The reality is that these platforms are media firms[56] and governing entities,[57] as well as advertising companies.[58] And while these platforms are media companies, they have a crucial distinction from traditional media companies.[59] Traditional media firms, such as news organizations, are

55. Balkin, *supra* note 33, at 1183; *see also* Alexis C. Madrigal, *The 'Platform' Excuse Is Dying*, ATLANTIC (June 11, 2019), https://www.theatlantic.com/technology/archive/2019/06/facebook-and-youtubes-platform-excuse-dying/591466 [https://perma.cc/BVX5-X2KN] (arguing that platforms like YouTube and Facebook have long absconded any responsibility for the impact of their sites and their policies by claiming to be tools for free expression).

56. Balkin, *supra* note 33, at 1183 ("[M]any digital infrastructure providers gradually recognized that they were media companies, and that in many cases they were governing communities of end-users, whether they liked it or not."). Technology companies reject any designation as a media company. *See, e.g.*, Jessica Guynn, *Zuckerberg: Facebook Isn't a 'Traditional' Media Company*, USA TODAY (Dec. 21, 2016), https://www.usatoday.com/story/tech/news/2016/12/21/mark-zuckerberg-facebook-not-a-traditional-media-company/95717102 [https://perma.cc/P8L3-89A3]. *But see* Alex Kantrowitz, *Twitter Embraces Its Role as a Media Company*, BUZZFEED NEWS (Dec. 28, 2016), https://www.buzzfeednews.com/article/alexkantrowitz/twitter-embraces-its-role-as-a-media-company [https://perma.cc/AZ6F-TGLL].

57. Balkin, *supra* note 33, at 1183. *See generally* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1602–03 (2018) (noting the similarities between private platforms and governance systems); Franklin Foer, *Facebook's War on Free Will*, GUARDIAN (Sept. 19, 2017), https://www.theguardian.com/technology/2017/sep/19/facebooks-war-on-free-will [https://perma.cc/D277-7REU] (quoting Mark Zuckerberg, C.E.O. of Facebook, saying: "In a lot of ways Facebook is more like a government than a traditional company").

58. *See* Dissenting Statement of Federal Trade Commissioner Rohit Chopra, *In re* Facebook, Inc., F.T.C. File No. 1823109, at 2 (July 24, 2019); Madrigal, *supra* note 55 ("'We're building,' Facebook's COO, Sheryl Sandberg, remarked, 'the world's first ad platform that delivers personalized marketing at scale.'"); Grafanaki, *supra* note 39, at 125 ("For the most part, the revenue of the platforms comes from advertising.").

59. Balkin, *supra* note 33, at 1192 ("These mass media companies were not conduits for the speech for the vast majority of the people who constituted the audience for their products. Rather, these companies (1) produced their own content, (2) published the content of a small number of creative artists, or (3) delivered content made by other organizations to a mass public. In the twentieth century model, the vast majority of people were members of an audience for mass media products, but very few actively used mass media as speakers or broadcasters. Twenty-first century governors of digital speech, by contrast, make their money by facilitating and encouraging the production of content by ordinary people and governing the communities of speakers that result. New media companies like Facebook, Google, YouTube, and Twitter do not produce most of the content they serve. Rather, their business model requires them to

conduits of speech for the select few[60] and produce the content in authorship and editorial roles. Platforms are conduits of speech for ordinary people and do not produce the content.[61] Instead of producing content for consumer purchase, platforms rely on advertising revenue and demand that as many users speak for as long, and as often, as possible.[62] This business model works, for platforms at least, by facilitating the constant production of user-developed content in order to capture the attention of other users.[63] This constant production of user-developed content and its subsequent consumption increases views and engagement with advertisements on the platform.[64] The more content on a user's Instagram newsfeed, for example, the more time that user will spend scrolling through the newsfeed and, thus, the more targeted ads they will view. Additionally, a user's online behavior and interactions with content, such as likes and comments, produces data, and, simultaneously, the content of user posts, such as geolocation or tags to other social media profiles, generates collectable user data.[65] All of this data enables more specific and voluminous targeted advertising on platforms.[66]

2. The Tension Between Platforms' Attention-Seeking Goal and Their Responsibility to Users

Platforms have a simple goal—shepherd users online and retain their attention.[67] Because most popular platforms' services are free, most of their revenue stems from data collection

induce as many people as possible around the world to post, speak, and broadcast to each other.").

60. Kate Vinton, *These 15 Billionaires Own America's News Media Companies*, FORBES (June 1, 2016), https://www.forbes.com/sites/katevinton/2016/06/01/these-15-billionaires-own-americas-news-media-companies/?sh=1ad558f660ad [https://perma.cc/XX57-RBCT].

61. Balkin, *supra* note 33, at 1192.

62. *Id.*

63. Gilad Edelman, *Social Media CEOs Can't Defend Their Business Model*, WIRED (Mar. 25, 2021), https://www.wired.com/story/social-media-ceo-hearing-cant-defend-business-model [https://perma.cc/LKC2-P8F6].

64. *Id.*

65. *See id.*

66. *Id.*

67. Grafanaki, *supra* note 39, at 125; Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard*, 1 J. FREE SPEECH L. 227, 263 (2021) ("A common critique of systems that amplify or target content based on user behavior is that, in conjunction with ads-based business models, they cause platforms to amplify anything that keeps users engaged, including harmful, misleading, or extreme content.").

and advertising.[68] Platforms' aim, then, is to keep users on their sites, making users provide their data and view advertisements. To keep users engaged, platforms are responsive to user demands,[69] or at least try to appear to be, such as user calls to moderate for misinformation.[70] Platforms' incentive to respond to users' demands, as well as those of third-party organizations representing user interests, catalyzes the creation and enforcement of user-identified norms for appropriate online behavior.[71] In addition to the creation and enforcement of these norms, platforms maintain user attention by using AI to predict which content a user might engage with.[72] Platforms can precisely predict user interest and feed them a personalized information diet because Big Data enables the tracking of each user's online activity and instantly analyzes user content consumption.[73] Then, the

68. *See* Keller, *supra* note 67, at 263; *see, e.g.*, *Meta's (Formerly Facebook Inc.) Advertising Revenue Worldwide from 2009 to 2021*, STATISTA (Feb. 2022), https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/#:~:text=In%202020%2C%20about%2097.9%20percent,increase%20in%20comparison%20to%20the [https://perma.cc/8EZ4-ZUUM] (stating that 97.9% of Facebook's revenue—a whopping $86 billion dollars—came from advertising).

69. Klonick, *supra* note 57, at 1627–30 (describing the economic reasons platforms moderate content pursuant to user norms).

70. Corin Faife & Dara Kerr, *Facebook Said It Would Stop Recommending Anti-Vaccine Groups. It Didn't*, MARKUP (May 20, 2021), https://themarkup.org/citizen-browser/2021/05/20/facebook-said-it-would-stop-recommending-anti-vaccine-groups-it-didnt [https://perma.cc/KJ5P-DD3W]; Sheera Frenkel, *Lies on Social Media Inflame Israeli-Palestinian Conflict*, N.Y. TIMES (May 18, 2021), https://www.nytimes.com/2021/05/14/technology/israel-palestine-misinformation-lies-social-media.html [https://perma.cc/DG9E-YUZ3].

71. Balkin, *supra* note 33, at 1183; *see, e.g.*, *Pinterest Embraces Body Acceptance with New Ad Policy*, PINTEREST NEWSROOM (July 1, 2021), https://newsroom.pinterest.com/en/post/pinterest-embraces-body-acceptance-with-new-ad-policy [https://perma.cc/Z44L-36R2] (describing Pinterest's collaboration with the National Eating Disorders Association to formulate a more body-positive advertising policy on its platform).

72. *See, e.g.*, Sam Biddle, *Facebook Uses Artificial Intelligence to Predict Your Future Actions for Advertisers, Says Confidential Document*, INTERCEPT (Apr. 13, 2018), https://theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai [https://perma.cc/G65M-YP48] (describing a confidential document detailing a new Facebook advertising strategy that uses AI to predict what users will purchase, how they will think, and how they might behave online).

73. Grafanaki, *supra* note 39, at 126 ("Big Data technologies now allow for precise tracking and analysis at the moment of content consumption. They also allow for instantaneous adjustment of the content selection based on the feedback. Traditional *editors* never had that kind of power.") (emphasis in original).

platform's content curation is automatically adjusted for that user.[74]

Platforms have a normative responsibility to their users—to moderate and ensure the safety of the content on their sites.[75] Platforms moderate content in various ways. Moderation can occur either before content is published or after content has been published and viewed by the masses.[76] Some content moderation is performed manually by humans, but most content moderation is automatically conducted by software.[77] Absent a legal obligation to moderate content,[78] there are two main reasons why platforms are compelled to moderate content.

First, much of platforms' motivation to regulate the speech and activity on their sites flows from their capitalistic goal to keep users satisfied with the platform[79]—or at least compulsively scrolling their feed.[80] Many platforms' content moderation policies rely on community standards, generated based on the social norms and values of each user.[81] Just as AI draws from users' online behavior to curate an engaging user experience, AI

---

74. *Id.*

75. *See, e.g.*, Biddle, *supra* note 72 (quoting Mark Zuckerberg, CEO of Facebook, telling lawmakers: "Across the board, we have a responsibility to not just build tools, but to make sure those tools are used for good").

76. Klonick, *supra* note 57, at 1635.

77. *Id.*

78. *See infra* Part I.B.

79. Klonick, *supra* note 57, at 1615, 1630; *see, e.g.*, Jeff Horwitz, *Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt.*, WALL ST. J. (Sept. 13, 2021), https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353 [https://perma.cc/Y775-HQDL] (describing an internal Facebook document stating that angering influential users is "PR risky," and therefore a different standard of content moderation of their content was permissible); *see also* Nina I. Brown, *Regulatory Goldilocks: Finding the Just and Right Fit for Content Moderation on Social Platforms*, 8 TEX. A&M L. REV. 451, 455 (2021) ("Only after the most egregious abuses—and particularly following threatened legal action or loss of advertisers—have social media platforms responded by removing content, making (often minor) policy changes, deleting user accounts, or amending terms of service.").

80. Allison Zakon, Note, *Optimized for Addiction: Extending Products Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 2020 WIS. L. REV. 1107, 1114; Angela Watercutter, *Doomscrolling Is Slowly Eroding Your Mental Health*, WIRED (June 25, 2020), https://www.wired.com/story/stop-doomscrolling [https://perma.cc/M29E-UXER].

81. Klonick, *supra* note 57, at 1632.

also moderates user-developed content.[82] A second influence on platforms' content moderation comes from the government.[83] Each platform ostensibly implements its own procedures and policies to moderate user-developed content, but they are influenced by direct requests from governments to comply with local laws or by less direct lobbying efforts.[84] Regardless of the underlying reason for their adoption, platforms' content moderation practices play an enormous role in shaping online public discourse.

3. Online Speech and Offline Impact

There is no overstating the impact of the internet. Even in its simplest form, it was revolutionary.[85] The modern internet and its platforms have hosted revolutions against oppressive regimes and white supremacy[86] and yet, the internet can be a tough place to be.[87] The internet's sometimes pernicious impact is not bound to the web. The introduction of anonymous posting, the amplification of user-developed content that painfully strikes the right nerve, and the creation of precise echo chambers

82. *Id.* at 1632–53 (describing YouTube and Facebook's evolution from moderating content based on community standards to codifying those standards into rules).

83. *Id.* at 1650–52 (describing platforms' geoblocking to conform to different countries' laws, and platforms cooperating with government requests, such as law enforcement requests for information).

84. *Id.* at 1650.

85. *See* Zaryn Dentzel, *How the Internet Has Changed Everyday Life*, OPENMIND (2013), https://www.bbvaopenmind.com/en/articles/internet-changed-everyday-life [https://perma.cc/X5P4-HYBL] (describing the myriad of ways the internet has changed everyday life, from the way people connect and consumer purchasing habits, to information consumption and sharing).

86. *See, e.g.*, Bijan Stephen, *Social Media Helps Black Lives Matter Fight the Power*, WIRED (Nov. 2015), https://www.wired.com/2015/10/how-black-lives-matter-uses-social-media-to-fight-the-power [https://perma.cc/QF9P-8GVT] (describing the role of social media platforms to protest police killings of Black and Brown people); Peter Beaumont, *The Truth About Twitter, Facebook and the Uprisings in the Arab World*, GUARDIAN (Feb. 25, 2011), https://www.theguardian.com/world/2011/feb/25/twitter-facebook-uprisings-arab-libya [https://perma.cc/G6CC-N6QT] (describing the use of internet platforms to organize against oppressive regimes in Libya, Tunisia, and Egypt).

87. Keller, *supra* note 67, at 257 ("People on the internet are terrible.").

have all permitted and exacerbated the discrimination and harassment of marginalized groups on a colossal scale.[88]

The phenomenon of "online disinhibition" was coined by psychologist Dr. John Suler to describe the way internet users first, shed their identity, and second, release behavioral constraints and act out online.[89] Unsurprisingly, certain identities are more susceptible to cyber-harassment than others. Young women are reported to have experienced online sexual harassment at a rate of 33%, compared to 11% for men.[90] The rate of sexual harassment against non-binary people is unfortunately not reported. Seven in ten LGBTQIA+ internet users report experiencing online harassment, with 51% reporting more severe forms of online abuse such as stalking and physical threats.[91] This is compared to a rate of four in ten heterosexual internet users who report having experienced online harassment.[92] Of the 41% of U.S. adults that reported experiencing online harassment, 54% of Black users and 47% of Hispanic users believed the harassment was race-based.[93]

To maintain user engagement, platforms have used AI to identify and amplify content that is "harmful, misleading, or extreme."[94] Users are more likely to engage with this type of content because it is divisive, sensational, and motivates them to

88. Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 68.

89. John Suler, *The Online Disinhibition Effect*, 7 CYBERPYSCH. & BEHAV. 321, 321 (2004).

90. Emily A. Vogels, *The State of Online Harassment*, PEW RSCH. CTR. (Jan. 13, 2021), https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment [https://perma.cc/22QM-Y4HS].

91. *Id.*; *see also* Alex Hern, *Young LGBTQ+ People More than Twice as Likely to Experience Hate Speech Online*, GUARDIAN (Oct. 17, 2022), https://www.theguardian.com/society/2022/oct/17/young-lgbtq-people-more-than-twice-as-likely-to-experience-hate-speech-online [https://perma.cc/8T9K-RNA5] (citing research that LGBTQIA+ young people are more than twice as likely to experience online hate speech than those who identify as heterosexual).

92. Vogels, *supra* note 90.

93. *Id.*

94. Keller, *supra* note 67, at 263; *see, e.g.*, The Journal, *The Facebook Files, Part 4: The Outrage Algorithm*, WALL ST. J. (Sept. 18, 2021) [hereinafter *The Facebook Files*], https://www.wsj.com/podcasts/the-journal/the-facebook-files-part-4-the-outrage-algorithm/e619fbb7-43b0-485b-877f-18a98ffa773f [https://perma.cc/T25R-TRAT] (describing Facebook's implementation of an algorithm known to exacerbate user anger and user engagement).

share their reaction.[95] In the words of Professor Daphne Keller, former Associate General Counsel for Google:

> At a societal level, they have spread misleading political material, to the detriment of democratic governance. At an individual level, they may lead dieters to content promoting anorexia, or viewers of Trump rallies to videos denying the Holocaust. Facebook's friend- and group-recommendation algorithms are said to have brought together violent right-wing extremists, one of whom ultimately shot and killed two people in Kenosha, Wisconsin.[96]

Numerous studies link social media use to poor mental health,[97] and many tech experts have acknowledged the collective harm of the internet.[98] These experts predict the internet will hurt democracy, if it hasn't already,[99] basing this prediction on the speed and scope of reality distortion occurring online,[100]

---

95. *The Facebook Files*, *supra* note 94.

96. Keller, *supra* note 67, at 230 (citations omitted).

97. *See* Luca Braghieri, Ro'ee Levy & Alexey Makarin, *Social Media and Mental Health*, 112 AM. ECON. REV. 3660, 3689 (2022) (finding evidence that the introduction of Facebook on college campuses was linked to increased levels of depression, anxiety, and poor mental health symptoms); Fazida Karim, Azeezat A. Oyewande, Lamis F. Abdalla, Reem Chaudhry Ehsanullah & Safeera Khan, *Social Media Use and Its Connection to Mental Health: A Systematic Review*, CUREUS 1, 5 (June 15, 2020), https://www.cureus.com/articles/31508-social-media-use-and-its-connection-to-mental-health-a-systematic-review [https://perma.cc/QA6D-AXW5] (finding that social media use is often found to have a detrimental effect on its users' mental health, particularly through social media envy and its consequential impact on user anxiety and depression); *cf.* Tammy Qiu, *A Psychiatrist's Perspective on Social Media Algorithms and Mental Health*, STAN. U. HUM.-CENTERED A.I. (Sept. 14, 2021), https://hai.stanford.edu/news/psychiatrists-perspective-social-media-algorithms-and-mental-health [https://perma.cc/W774-BEK8] (describing one psychiatrist's proposal for reimagining more empathetic social media).

98. Janna Anderson & Lee Rainie, *Many Tech Experts Say Digital Disruption Will Hurt Democracy*, PEW RSCH. CTR. (Feb. 21, 2020), https://www.pewresearch.org/internet/2020/02/21/many-tech-experts-say-digital-disruption-will-hurt-democracy [https://perma.cc/9XPL-SMP8].

99. Adrienne LaFrance, *"History Will Not Judge Us Kindly*," ATLANTIC (Oct. 25, 2021), https://www.theatlantic.com/ideas/archive/2021/10/facebook-papers-democracy-election-zuckerberg/620478 [https://perma.cc/5YQ9-6V7H] (citing Facebook's internal research determination that the platform was actively harming democracy).

100. Tomas Chamorro-Premuzic, *How the Web Distorts Reality and Impairs Our Judgment Skills*, GUARDIAN (May 13, 2014), https://www.theguardian.com/media-network/media-network-blog/2014/may/13/internet-confirmation-bias [https://perma.cc/726C-DNNX].

the impact of surveillance capitalism,[101] and the decline of journalism in the digital age.[102]

Algorithms and AI are integral components of online platforms, simultaneously maintaining user engagement and enabling large-scale content moderation of online speech. There is an inherent tension between platforms' goal to increase and sustain user engagement and their normative responsibility to ensure their platforms are not harmful to users. This tension is realized in the offline impacts of online speech.

B. AS CDA 230'S IMMUNITY EXPANDS, SO DOES THE INTERNET'S REACH, RESULTING IN MORE ONLINE SPEECH, GREATER IMPACT, AND FEWER CONSEQUENCES

Before analyzing prospective solutions to online speech regulation, it is important to first understand the law that "created the [i]nternet"[103] and the "mighty fortress" surrounding it.[104] This Part describes CDA 230 at its enactment and as it exists today, arguing that CDA 230's blanket immunity has largely stayed firm as platforms have grown more autonomous in their dissemination of user-developed content.

1. Congressional Intent

In 1994, an anonymous user posted to Prodigy, a subscription-based platform that hosted bulletin boards, news, weather, email, and sports reporting.[105] The user alleged that the head of a Long Island securities brokerage firm, Stratton Oakmont, had committed fraudulent and criminal acts.[106] Stratton Oakmont filed suit against Prodigy claiming that the platform was acting like a publisher and, thus, was liable for defamation.[107] The

101. John Laidler, *High Tech Is Watching You*, HARV. GAZETTE (Mar. 4, 2019), https://news.harvard.edu/gazette/story/2019/03/harvard-professor-says-surveillance-capitalism-is-undermining-democracy [https://perma.cc/YB4X-KPXY] (interviewing Shoshana Zuboff, author of *The Age of Surveillance Capitalism*, who said: "I define surveillance capitalism as the unilateral claiming of private human experience as free raw material for translation into behavioral data.").

102. Anderson & Rainie, *supra* note 98.

103. KOSSEFF, *supra* note 22.

104. Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 406 (2017).

105. Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995).

106. *Id.*

107. *Id.* at *2.

court sided with the brokerage firm, deciding that Prodigy's implementation of content guidelines and screening for offensive postings made the platform a publisher.[108] As a publisher, Prodigy was therefore liable for the defamatory remarks contained within the user-developed content on its site.[109] Essentially, the court held that because Prodigy imperfectly attempted to moderate its platform and monitor user-developed content, Prodigy was liable for *any* offensive posts that evaded its monitoring system.[110] The *Prodigy* decision also meant that, counterintuitively, if a platform did not attempt to moderate user-developed content at all, the platform escaped liability.[111] Conversely, under *Prodigy*, if a platform attempted to moderate content posted on its site, but unlawful content mistakenly slipped through, the platform would be liable for any legal repercussions arising from the content.[112]

Congress understood that *Prodigy* could disincentivize websites from regulating user-developed content and enable a proliferation of obscene content online. When the district court found Prodigy liable for its inexact content moderation, Congress passed a slight but impactful provision into the Communications Decency Act—Section 230.[113]

CDA 230 aimed to overturn *Prodigy*, but Congress was also explicit about its vision of unfettered innovation on the internet.[114] Congressman Bob Goodlatte, one of CDA 230's co-spon-

---

108. *Id.* at *4 ("Prodigy implemented [control over content] through its automatic software screening program, and the Guidelines which Board Leaders are required to enforce. . . . Prodigy is clearly making decisions as to content, and such decisions constitute editorial content.") (citations omitted).

109. *Id.*

110. *Id.* ("Prodigy's conscious choice, to gain the benefits of editorial control, has opened it up to a greater liability than . . . other computer networks that make no such choice.").

111. Jeff Kosseff, *Defending Section 230: The Value of Intermediary Immunity*, 15 J. TECH. L. & POL'Y 123, 129 (2010).

112. *Id.*

113. H.R. REP. NO. 104-458, at 194 (1996) (Conf. Rep.) ("One of the specific purposes of this section is to overrule *Stratton-Oakmont v. Prodigy* and any other similar decisions which have treated such providers and users as publishers or speakers of content that is not their own because they have restricted access to objectionable material.").

114. 141 CONG. REC. 22,022 (1995) ("It is the policy of the United States to—(1) promote the continued development of the Internet and other interactive computer services and other interactive media; (2) preserve the vibrant and

sors, explained that the internet, as an emerging industry, required few constraints: "We are talking about something that is far larger than our daily newspaper. We are talking about something that is going to be thousands of pages of information every day, and to have that imposition [to edit that information] imposed on them is wrong."[115] Recognizing the impossible burden that mandating perfect content moderation would force on the fledgling internet industry, the following two-part objective was enshrined and enacted in the following statutory language:

> (c) Protection for "Good Samaritan" blocking and screening of offensive material
>
> (1) Treatment of publisher or speaker.
>
> No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.
>
> (2) Civil liability.
>
> No provider or user of an interactive computer service shall be held liable on account of—
>
> (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene lewd, lascivious, filthy, excessively violent, harassing, or otherwise, objectionable, whether or not such material is constitutionally protected; or
>
> (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).[116]

This provision first protected websites from liability as the "publisher" or "speaker" of the user-developed content posted on its site. The provision also immunized websites from liability for any action or inaction to remove user-developed content. In theory and intent, CDA 230 incentivized websites to self-regulate user-developed content by granting immunity to websites from any liability for user-developed content. In practice, it is debatable whether CDA 230 incentivized platforms' regulation of user-developed content.[117]

competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation.").

115. *Id.* at 22,046.

116. 47 U.S.C. § 230(c).

117. *See* Checkstep, *The Evolution of Content Moderation Rules Throughout the Years*, MEDIUM (Apr. 1, 2021), https://medium.com/checkstep/the-evolution-of-content-moderation-rules-throughout-the-years-bccc9859cb31 [https://perma.cc/ZP5V-F25K] (arguing that social media platforms are caught between the need to "curb[] the rise of hate speech [and] disinformation," and the need to also "protect the fundamental right of expression for their users").

2. Judicial Interpretation Stretches the CDA 230 Shield and Congressional Intent Is Lost

Since 1996, without meaningful direction from the Supreme Court, courts have broadened the scope and impact of CDA 230 to immunize platforms from liability for algorithmic actions. AI and its impact on user-developed content, however, were not considerations in the hearings and drafting culminating in CDA 230's enactment. This Note argues that there has been such a drastic online evolution, an evolution for which algorithms and AI have proven pivotal, that it is necessary to re-imagine a regulatory scheme for online speech, and this scheme must consider the role and impact of algorithms and AI on online speech.

a. The Confirmation of Internet Exceptionalism

Because the Supreme Court has remained practically silent on the matter,[118] the Fourth Circuit, interpreting CDA 230 just one year after its passage, set the tone for its judicial interpretation. In *Zeran v. American Online, Inc.*,[119] American Online, Inc., commonly known as "AOL," was sued for its alleged delay in removing defamatory user-developed content from its platform.[120] The Fourth Circuit, focusing on the congressional objective to establish a safe harbor for internet innovation,[121] held that CDA 230 barred any cause of action that would hold "service providers

118. *But see* Malwarebytes, Inc. v. Enigma Software Grp. USA, LLC, 141 S. Ct. 13, 13 (2020) (Thomas, J., statement regarding denial of certiorari) (arguing that CDA 230 interpretation has been broadened beyond the plain meaning or statutory intent of Congress); Biden v. Knight First Amend. Inst. at Columbia Univ., 141 S. Ct. 1220, 1221 (2021) (Thomas, J., concurring) ("I write separately to note that this petition highlights the principal legal difficulty that surrounds digital platforms—namely, that applying old doctrines to new digital platforms is rarely straightforward."). In October 2022, the Supreme Court granted certiorari to hear arguments from a Ninth Circuit case and may ultimately decide whether CDA 230 immunizes YouTube and similar websites that make recommendations of user-developed content, or if CDA 230 only limits liability when websites engage in traditional editorial functions, such as content moderation decisions. *Gonzalez v. Google LLC*, SCOTUSBLOG, https://www.scotusblog.com/case-files/cases/gonzalez-v-google-llc [https://perma.cc/8958-YQPF].

119. 129 F.3d 327 (4th Cir. 1997).

120. *Id.* at 328.

121. *Id.* at 330 ("[Congress] found that the Internet and interactive computer services 'have flourished, to the benefit of all Americans, *with a minimum of government regulation*.' Congress further stated that it is 'the policy of the United States . . . to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, *unfettered by Federal or State regulation*.'") (citations omitted) (emphasis in original).

liable for information originating with a third-party user."[122] In prioritizing unfettered free speech on the internet, the *Zeran* court narrowly focused on one legislative purpose of CDA 230.[123] This decision had a profound impact on subsequent judicial interpretations of CDA 230 and "built a mighty fortress protecting platforms from accountability for unlawful activity on their systems."[124] As CDA 230 has been interpreted, this slight provision shields platforms from liability for surprising actions, including the republishing of illegal content,[125] encouraging internet users to post unlawful content,[126] altering its own sitewide policies to buoy illegal activity,[127] and selling and advertising dangerous or defective products.[128]

Zeran confirmed an attitude of Internet Exceptionalism, the idea alluded to by Congressman Goodlatte, that the internet was so fragile and unique that it could not be subject to traditional forms of legal accountability.[129] And the 1997 decision might

---

122. *Id.*

123. Citron & Wittes, *supra* note 104, at 408. In its opinion, the *Zeran* Court insisted that the CDA was enacted solely to reflect "Congress' desire to promote unfettered speech on the [i]nternet." *Zeran*, 129 F.3d at 334. *Zeran* ignored the text and the statutory history to protect "Good Samaritan" action to regulate offensive material. Citron & Wittes, *supra* note 104, at 407–08.

124. Citron & Wittes, *supra* note 104, at 406.

125. *See, e.g.*, Phan v. Pham, 105 Cal. Rptr. 3d 791, 795 (Ct. App. 2010) (extending CDA 230 immunity to an internet intermediary who forwarded a defamatory email and added its own message of "[e]verything will come out to the daylight").

126. S.C. v. Dirty World, No. 11-CV-00392-DW, 2012 WL 3335284, at *4–5 (W.D. Mo. Mar. 12, 2012) (holding that CDA 230 immunity applied for a gossip website that only published select submissions and includes a function where users can get post feedback from the site operator).

127. Doe v. Backpage.com, LLC, 817 F.3d 12, 21–22 (1st Cir. 2016) (holding that the operator of Backpage, an online classified advertisement forum, was immune from liability for third-party use of the cite to facilitate sex trafficking).

128. Hinton v. Amazon.com.dedc, LLC, 72 F. Supp. 3d 685, 692 (S.D. Miss. 2014) (dismissing a case brought against eBay for advertising and selling merchandise it knew had been recalled).

129. Neil Fried, *The Myth of Internet Exceptionalism: Bringing CDA 230 into the Real World*, AM. AFFS. J. (May 20, 2021), https://americanaffairsjournal.org/2021/05/the-myth-of-Internet-exceptionalism-bringing-section-230-into-the-real-world [https://perma.cc/V9UR-GK5N]; *see also* Ashley Deeks, *Facebook Unbound?, Foreword* to 105 VA. L. REV. ONLINE 1, 6–7 (2019) ("First, members of Congress lack sophisticated understandings of how these companies—and the technologies that undergird their products—work. . . . Second, knowing what to regulate, in what level of detail, and at what stage in the overall devel-

have been reasonable in its temporal and social context as it was issued only a few years after the internet reached the mainstream. The internet, however, is no longer exceptional.[130] Rather, the internet is a ubiquitous force that cannot be disentangled from today's societal infrastructure. Even Professor Jeff Kosseff, noted internet legal scholar and longtime CDA 230 advocate, has conceded that the law is due an amendment to account for new technologies and to avoid absurd outcomes.[131]

Because congressional intent behind CDA 230 was to incentivize platforms to self-regulate user-developed content, and because Congress was careful not to stunt the internet's growth, the courts have stretched CDA 230's blanket immunity too far and enabled platforms to grow comfortable to the point of complicity with harm to users.

II. OLD SCHOOL V. NEW SCHOOL: ONLINE SPEECH REGULATION DISTINCTION

It is a well-established communications principle that "the medium is the message."[132] Put differently, it is not only the substance of a message, but also how it is relayed, that is impactful to how the communication is received.[133] Platforms have become a popular communication medium, conveying massive quantities of speech by enabling users to post their own user-developed content. As a medium, platforms have a unique ability to amplify and influence the reception of the user-developed content. This

opment of technologies such as machine learning is simply hard. . . . Third, Congress fears undercutting U.S. innovation by regulating too soon.").

130. *See, e.g.*, Lauren Easton, *Ready to Lowercase "Internet" and "Web*,*"* ASSOCIATED PRESS (Apr. 2, 2016), https://blog.ap.org/products-and-services/ready-to-lowercase-internet-and-web [https://perma.cc/GGE4-WC74] (reporting the 2016 AP Stylebook would lowercase the words "internet" and "web"); CHICAGO MANUAL OF STYLE § 7.80 (U. Chi. Press ed., 17th ed. 2017) (lowercasing "internet").

131. KOSSEFF, *supra* note 22 ("Congress passed Section 230 at the peak of the Internet exceptionalism era. Twenty years later, the Internet is less exceptional. It is not a cool new technology. . . . The Internet is not as exceptional as it was in 1996 because it is now woven into the fabric of nearly every aspect of life. It is also more complicated, with artificial intelligence and complex algorithms processing third-party content while employing power and capabilities that probably were not contemplated in 1996. . . . [W]e should all work to understand how to improve Section 230.").

132. MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN 23–35 (1964).

133. *Id.*

medium, however, has not been subject to the law.[134] Professor Jack M. Balkin offers a simple framework to understand a complicated topic—government regulation of online speech. Balkin describes the traditional speech regulation framework pre-dating the internet, which he aptly calls Old School Speech Regulation. The government's Old School Speech Regulation is directed at people, spaces, and the predigital technologies of mass distribution, such as newspapers and magazines.[135] Conversely, and in addition to Old School Speech Regulation, the twenty-first century requires an emerging New School Speech Regulation. New School Speech Regulation is aimed at the digital infrastructure that conveys the substantive messages of online speech, rather than the speaker or publisher of the message.[136]

Balkin's framework illustrates one explanation for why CDA 230 is unable to regulate online speech. Because New School Speech Regulation targets the digital infrastructure buttressing online speech, CDA 230, a law that immunizes platforms from liability for that speech, renders New School Speech Regulation impossible.

A. UNDERSTANDING OLD SCHOOL SPEECH REGULATION

Balkin's Old School Speech Regulation is the legal framework traditionally studied in First Amendment law school courses—to regulate speech, the government directly regulates "speakers" and "publishers."[137] The two watershed cases that founded Old School Speech Regulation are *New York Times Co. v. Sullivan*[138] and *New York Times Co. v. United States*[139] *(Pentagon Papers).* Together, these cases solidified the legal truths that a democratic society requires a healthy press and that robust public discourse is vital to democratic legitimacy.[140] Using traditional enforcement methods such as fines, injunctions, and imprisonment, the judiciary of Old School Speech Regulation

134. *See supra* Part I.A, I.B.

135. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2306 (2014).

136. *Id.*

137. *Id.* at 2308.

138. 376 U.S. 254, 283 (1964) (holding the First Amendment protects the freedom of the press against claims brought by a public official).

139. 403 U.S. 713, 714 (1971) (per curiam) (holding the First Amendment protected the newspaper's right to publish classified government documents in the public interest).

140. Balkin, *supra* note 135, at 2296.

strives to maintain an informed press and uninhibited public discourse.[141] The Old School Speech Regulation paradigm is dyadic, with the state on one side and publishers and speakers on the other.[142] Old School Speech Regulation is conceptually simple—the government targets A to regulate the speech of A. In both *Sullivan* and *Pentagon Papers*, for example, the government unsuccessfully targeted the New York Times as the publisher of the messages at issue. The Old School Speech Regulation framework has been effectively employed to regulate most forms of speech, but, as the mediums of speech have changed, a new speech regulation framework has emerged.

B. NEW SCHOOL SPEECH REGULATION

For purposes of this Note, New School Speech Regulation can be simply understood as the government, C, targeting platforms, B, to regulate the speech of user A. The modern internet has disrupted the Old School Speech Regulation paradigm by powerfully introducing a third class of actors—the private companies that create and maintain digital infrastructure.[143] The government regulates online speech by aiming at the speech of platform users.[144] Under the New School Speech Regulation paradigm, the government does such regulating by coercion or compulsion.[145] New School Speech Regulation forms a triangle with nation states, digital infrastructure companies, and ordinary people communicating through the digital infrastructure at each point.[146]

Importantly, CDA 230 has prevented the development of First Amendment doctrine concerning platforms' moderation of

141. *Id.*

142. Balkin, *supra* note 33.

143. Balkin, *supra* note 135, at 2298.

144. Balkin, *supra* note 33, at 1180 ("Sometimes, as in the case of digital surveillance and threats to cybersecurity, states and infrastructure operators have common concerns. But equally often, companies are pushed, cajoled, and coerced into cooperation—then they are effectively coopted into assisting states in governance.").

145. Genevieve Lakier, *Informal Government Coercion and the Problem of "Jawboning*,*"* LAWFARE (July 26, 2021), https://www.lawfareblog.com/informal-government-coercion-and-problem-jawboning [https://perma.cc/95YU-EQC6].

146. Balkin, *supra* note 33, at 1187–89. Balkin is clear, however, that because of the complexity of the modern internet and its myriad of compounding compositional forces, this new speech regime is not a triad. *Id.* at 1191–93.

user-developed content.[147] Some scholars have argued that at least some aspects of CDA 230 immunity may already be promised to platforms by the First Amendment.[148] Putting such arguments aside, CDA 230 is incompatible with the New School Speech Regulation paradigm because it effectively dismisses all claims against platforms, even those related to the platform's affirmative actions to amplify or recommend user-developed content. Legal recourse from the judiciary has long been a crucial way the government has informed the public and companies about their duties to one another. Because CDA 230 effectively bars the litigation of claims against platforms, the judiciary is unable to nudge companies to act in accordance with First Amendment principles.[149]

Because of CDA 230's expansive judicial interpretation and robust shield for platforms,[150] an example from overseas can best illustrate New School Speech Regulation. In 2014, the Court of Justice of the European Union (CJEU) required Google Spain to remove links to news articles about a Spanish citizen's social security debts from over a decade prior.[151] In *Google Spain v. Agencia Espaiola de Proteccion de Datos*, rather than require the newspapers to remove the articles, the CJEU required the digital infrastructure provider, Google Spain, to remove links to the information.[152] The court sought not to target the "publisher" of the message, the newspapers that published the content at issue, but the platform as the digital infrastructure medium amplifying the content.[153] This is the exact New School Speech Regulation paradigm described in Balkin's work—the government regulates speech by regulating the digital infrastructure hosting the

---

147. *See* Alan Z. Rozenshtein, *Silicon Valley's Speech: Technology Giants and the Deregulatory First Amendment*, 1 J. FREE SPEECH L. 337, 348–49 (2021).

148. *Id.* at 349; *see also* Eric Goldman, *Why Section 230 is Better Than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33, 34 (2019).

149. Goldman, *supra* note 148, at 44 ("[C]ourts much prefer to rely on statutory grounds like Section 230 instead of interpreting the First Amendment . . . .").

150. *See supra* Part I.B.2.a.

151. Case C-131/12, Google Spain SL v. Agencia Española de Proteccion de Datos, ECLI:EU:C:2014:317, paras. 9, 94 (May 13, 2014), https://curia.europa.eu/juris/document/document_print.jsf?mode=DOC&pageIndex=0&docid=152065&part=1&doclang=EN&text=&dir=&occ=first&cid=395712 [https://perma.cc/3FMF-2QHN].

152. *Id.*

153. *Id.*

content and making it accessible to users.[154] The rationale in *Google Spain* would not hold in the United States because, under CDA 230, platforms and websites like Google are not legally responsible for rendering accessible user-developed content on their sites. Due to this, courts do not have a pathway by which to regulate digital infrastructures in the New School Speech paradigm.

1. Why the Move to New School Speech Regulation?

Old School Speech Regulation is still a suitable regulation scheme for many forms of speech. The dyadic regime has and continues to effectively regulate speech in traditional media, such as newspapers, cable television, and broadcast. The modern internet, however, has reshaped and complicated *a lot* of speech.[155] Platform users within the digital infrastructure may be anonymous, outside the country, or not even human.[156] The target in the dyadic structure can be illusive in digital space. Conversely, the platforms of digital infrastructure are well-known, typically with U.S. headquarters, and money[157]—a lot of money.[158] More importantly, these companies can regulate and govern speech far beyond the capacity of government because, as private companies, they are not constrained by the First Amendment.[159] Due to the internet's evolution and the impact of platforms' content moderation practices, the New School Speech

154. Balkin, *supra* note 135, at 2298.

155. *See supra* Part I.A.

156. Adrienne LaFrance, *The Internet Is Mostly Bots*, ATLANTIC (Jan. 31, 2017), https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043 [https://perma.cc/EH7U-T6PF].

157. Shira Ovide, *What Big Tech's Riches Mean for Our Future*, N.Y. TIMES (Feb. 3, 2022), https://www.nytimes.com/2022/02/03/technology/big-tech-facebook-earnings.html [https://perma.cc/UUQ6-9WZB].

158. *See* Balkin, *supra* note 33, at 1175 ("[T]he large private enterprises that constitute the digital infrastructure have the technical and bureaucratic capacity to regulate and govern speech . . . .").

159. *See* Sarah Ludington, Lauren Smith & Christian Bale, *How Social Media Platforms Can Promote Compliance with the First Amendment*, N.Y.U. L. REV. F. (Sept. 30, 2022), https://www.nyulawreview.org/forum/2022/09/how-social-media-platforms-can-promote-compliance-with-the-first-amendment [https://perma.cc/YGP7-RG5V] ("As a private entity, Twitter, like other social media companies, is not bound by the First Amendment and may adopt its own content moderation policies."); Klonick, *supra* note 57, at 1662 (analyzing platforms within traditionally recognized First Amendment categories and ultimately arguing that no existing First Amendment category seems to "precisely meet the descriptive nature of what online platforms are").

Regulation scheme is the framework that can most effectively ground any future regulation of online speech.

Professor Kate Klonick has convincingly described platforms as "the New Governors" of online speech, consistent with Balkin's development of the Old School/New School Speech Regulation frameworks.[160] After surveying Facebook, Twitter, and YouTube's content moderation policies, Klonick concluded that these platforms' content moderation practices largely mirror fundamental aspects of the U.S. legal system[161] and prioritize First Amendment principles.[162] However, according to Klonick, attempting to conceptualize online platforms within First Amendment jurisprudence as debated and understood today "misses much of what is actually happening in these private spaces."[163] Platforms are unlike typical speech distributors because of their ability to dictate which user-developed content is amplified and which is taken down. Rather than understand platforms' content moderation practices from the perspective of town squares or broadcast journalism, firmly rooted in First Amendment jurisprudence, Klonick advocates for analyzing online speech from the perspective of private governance and self-regulation.[164] This view captures the nuance and power of private platforms' self-governance models and their contribution to offline democracy.

2. Why Is the Move a Problem?

Platforms should not be able to set content moderation policies without true accountability to the public or input from government. Something as essential to democracy as speech and the marketplace of ideas should not be dictated by private entities

160. Klonick, *supra* note 57, at 1662 ("Analyzing online platforms from the perspective of governance is both more descriptively accurate and more normatively useful in addressing the infrastructure of this ever-evolving private space.").

161. *Id.* at 1664 ("Procedurally, platform content-moderation systems have many similarities to a legal system.").

162. *See id.* at 1644–47 (comparing principles of American law to content moderation practices of social media platforms, ultimately concluding that "[c]ontent moderators act in a capacity very similar to that of judges").

163. *Id.* at 1662; *see also id.* at 1663 ("Perhaps most significantly, the idea of governance captures the power and scope these private platforms wield through their moderation systems and lends gravitas to their role in democratic culture.").

164. *See id.* ("[A]nalysis of online speech is best considered from the perspectives of private governance and self-regulation.").

with their own economic priorities as a substitute for democratically developed principles and regulatory standards.

Digital infrastructure companies are privately owned and biased platforms. These qualities do not make platforms inherently problematic as entities, as there is no legal requirement that platforms be unbiased.[165] But do these characteristics make platforms good governors? Absent transparent direction from government or public accountability, certainly not. Klonick has identified two main concerns with platforms' new governance: (1) loss of equal access to and participation in speech on these platforms; and (2) lack of direct accountability between platforms and their users.[166]

In her survey of Facebook, Twitter, and YouTube, Klonick noted that when these platforms make their own rules and procedures around content moderation, typically with very little transparency, preferential treatment is given to some users over others.[167] Take, for example, Facebook's XCheck program.[168] XCheck gave millions of high-profile Facebook users special treatment and even exempted these users from some content moderation rules.[169] This characteristic of self-regulated content moderation means that, even though platforms enable the reach of ordinary people's speech at unprecedented rates, they fail to

165. It is a common misconception that CDA 230 requires platforms to be unbiased. *See* Citron & Franks, *supra* note 88, at 62 ("Not only does Section 230 not require platforms to act neutrally vis-à-vis political viewpoints as state actors should, it urges exactly the opposite."); Mary Anne Franks, *How the Internet Unmakes Law*, 16 OHIO ST. TECH. L.J. 10, 13 (2020) ("A number of high-profile politicians have claimed . . . that the law requires online intermediaries to be 'neutral platforms' or lose their immunity, a claim unsupported by the text of the statute or case law interpreting it."). In his defense of Parler, C.E.O. George Farmer boasted that the platform is a "neutral town square" whereas other platforms, such as Facebook, are inherently slanted and political. Sway, *supra* note 15. This misconception is twofold: (1) CDA 230(b)(4) states as one of the statute's purposes the removal of disincentives to the development of content moderation technologies; and (2) the First Amendment does not obligate the speech of private actors to be neutral as it does the speech of government actors. Citron & Franks, *supra* note 88, at 62.

166. Klonick, *supra* note 57, at 1664.

167. *See id.* at 1665 ("[P]rivate platforms are increasingly making their own choices around content moderation that give preferential treatment to some users over others.").

168. *See* Horwitz, *supra* note 79 (describing the preferential treatment received on Facebook regarding content moderation by certain "VIP users" through XCheck).

169. *Id.*

provide a "fair opportunity to participate."[170] Private governance of platforms is comparable to the ownership of media by the wealthy elite, and government regulation has played an important role in ensuring equal participation in the media.[171] While critiquing platforms' content moderation practices, Klonick proposes that future regulation of user-developed content carefully consider the platform governance structures already in place and the motivations behind them.[172]

Platforms' lack of transparency about decisions that directly impact users and indirectly shape society is particularly troubling because these platforms render users powerless.[173] Platforms are incredibly ingrained in the digital infrastructure, as well as social fabric of offline life, yet users are subject to the whims of the private corporations operating them.[174] Any user-derived power to shape these platforms and their content moderation practices is largely indirect, as exercised when users decrease time spent on the platforms and, subsequently, declining advertising views, rather than direct market empowerment.[175] Worse, because these platforms are built by AI with the goal of maintaining user engagement, platforms create an antidemo-

170. Klonick, *supra* note 57, at 1665 ("[T]he open acknowledgement of different treatment and rule sets for powerful users over others reveal that a fair opportunity to participate is not currently a prioritized part of platform moderation systems.").

171. *See id.* ("[T]hese problems are nothing new—they are quite similar to the concerns to democracy posed by a mass media captured by a powerful, wealthy elite." (citing Balkin, *supra* note 25, at 30)).

172. *See id.* at 1666 ("[A]ny proposed regulation—be it entirely new laws or modest changes to § 230—should look carefully at how and why the New Governors *actually* moderate speech. Such, if any, regulation should work with an understanding of the intricate self-regulatory structure already in place in order to be the most effective for users.") (citation omitted).

173. *See id.* ("[T]he central difficulty in simply allowing these systems to self-regulate in a way that takes into account the values and rights of their users is that it leaves users essentially powerless.").

174. *See id. But see id.* at 1627–30 (describing platforms' consideration of user norms and demands to remain economically viable).

175. *See id.* at 1666 ("[P]latforms are beholden to their corporate values, to the foundational norms of American free speech, and to creating a platform where users will want to engage. Only the last of these three motivations for moderating content gives the user any 'power,' and then only in an indirect and amorphous way.").

cratic space—an echo-chamber—of like ideas and catalyze "nondeliberative polarization."[176]

Understanding Balkin's New School/Old School Speech Regulation regimes provides a theoretical framework to consider the problems with the current interpretation of CDA 230. This discussion highlights the need for a modernized regulatory regime that understands and considers the impact of algorithms and AI on the internet.

III. POWER WITHOUT RESPONSIBILITY:[177] CDA 230'S LIMITATIONS IN THE ALGORITHMIC SOCIETY

In a transformed digital landscape, CDA 230 is an obstacle to addressing harms caused by online speech.[178] Because the statute was enacted with Old School Speech Regulation in mind—as opposed to the New School Speech Regulation that modern society requires—and the courts have overextended its shield to create a "mighty fortress,"[179] there is no existing online regulation that meets the challenges of the modern internet.

A. A MAJOR INSUFFICIENCY OF CDA 230: EMPLOYING THE OLD SCHOOL SPEECH REGULATION PARADIGM

CDA 230 was enacted and has been interpreted under the dyadic model of speech regulation. Old School Speech Regulation, however, cannot address online speech because of the introduction of a third actor—online platforms. This has justifiably

---

176. *Id.* at 1667 ("[T]he so-called echo-chamber effect, which creates an antidemocratic space in which people are shown things with which they already associate and agree, lead[s] to nondeliberative polarization."); *see* DeVos, *supra* note 52 (describing the echo chamber effect of social media).

177. Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 1002–15 (2008) (analyzing how platforms can set and moderate their own content standards, often against their user's wishes, while avoiding liability).

178. CDA 230 has been largely untouched since its 1996 enactment. *But see* Allow States and Victims to Fight Online Sex Trafficking (FOSTA) Act of 2017, Pub. L. No. 115-164, § 4, 132 Stat. 1253, 1254 (2018) (clarifying that CDA 230 does not provide legal protection to websites that facilitate sex trafficking and enhancing criminal penalties for persons who use websites to facilitate sex trafficking).

179. Citron & Wittes, *supra* note 104, at 406 ("[C]ourts have built a mighty fortress protecting platforms from accountability for unlawful activity on their systems—even when they actively encourage such activity or intentionally refuse to address it.").

left courts puzzled about how to square platforms' AI and algorithmic action within CDA 230 immunity, with the majority of courts deciding platforms' algorithmic action is immune under the statute. Such findings only exacerbate the judiciary's overextension of CDA 230 immunity, however.

One of the modern internet's defining characteristics is the use of algorithms and AI to build a digital infrastructure that shapes users' information diet and increases user engagement on platforms.[180] Courts have attempted to square CDA 230 in the context of the Algorithmic Society, with inconsistent and largely overexpansive results. The seminal federal decision addressing whether a platform's algorithmic action is subject to CDA 230 immunity is the Second Circuit's 2019 decision in *Force v. Facebook*.[181] In *Force*, family members of Americans killed by international terrorist organization Hamas sued Facebook, claiming that the platform's suggested friend algorithm, which was used by Hamas to recruit new members, was affirmative action, taken by the platform, beyond the scope of CDA 230 immunity, and that Facebook was thus liable for "providing material support" to Hamas.[182] The *Force* majority found that Facebook was not liable because the company was simply arranging and distributing user-developed information as an editor is known to do.[183] Although Chief Judge Katzmann agreed with the majority opinion with respect to foreign law claims and the extraterritorial application of CDA 230, he partially dissented to argue that Facebook's algorithm was not a traditional editorial function at all.[184] Rather, Katzmann continued that, by suggesting its users connect as "friends," Facebook created its own message to users by implying the users would like each

180. *See supra* Part I.A.

181. 934 F.3d 53 (2d Cir. 2019).

182. *Id.* at 57–61.

183. *See id.* at 66 ("[A]rranging and distributing third-party information inherently forms 'connections' and 'matches' among speakers, content, and viewers of content, whether in interactive internet forums or in more traditional media. That is an essential result of publishing.").

184. *Id.* at 82 (Katzmann, C.J., concurring in part and dissenting in part) ("[C]laims based on these algorithms do not inherently treat Facebook as the publisher of third-party content.").

other's content.[185] It was this algorithm-derived message, he argued, that contributed to offline social networks with consequences for the offline world.[186]

A few months after the *Force* ruling, the Ninth Circuit issued a decision on a similar legal question in *Dyroff v. Ultimate Software Group, Inc.*[187] In *Dyroff*, the plaintiff sued the operator of Experience Project, an anonymous social networking site, where Wesley Greer inquired about where to find heroin.[188] Through the site, Greer obtained heroin, which was laced with fentanyl and ultimately killed him.[189] Greer's mother brought the case against the Experience Project for allowing the trafficking of deadly narcotics, steering users to groups where these sales were facilitated, and sending alerts to posts within these groups.[190] Just as the Second Circuit did before it, the Ninth Circuit dismissed the case, finding that the site's recommendation algorithm was within CDA 230 immunity.[191]

Then in 2021, the Supreme Court of Texas issued an opposite ruling concerning Facebook's CDA 230 immunity for allegedly facilitating online connections.[192] The court held that Facebook could not be held liable under CDA 230 for failing to warn users of the risk of human trafficking on its platforms.[193] However, the court decided that Facebook could be held liable under a Texas human trafficking law for "affirmatively acting" to

---

185. *Id.* ("If a third party got access to Facebook users' data, analyzed it using a proprietary algorithm, and sent its own messages to Facebook users suggesting that people become friends or attend one another's events, the third party would not be protected as 'the publisher' of the users' information.").

186. *Id.*

187. 934 F.3d 1093 (9th Cir. 2019).

188. *Id.* at 1095.

189. *Id.*

190. *Id.*

191. *Id.* at 1098 ("By recommending user groups and sending email notifications, Ultimate Software, through its Experience Project website, was acting as a publisher of others' content. These functions—recommendations and notifications—are tools meant to facilitate the communication and content of others. They are not content in and of themselves.").

192. *In re* Facebook, Inc., 625 S.W.3d 80 (Tex. 2021).

193. *See id.* at 94 ("Like Plaintiffs' other common-law claims, these claims seek to hold Facebook liable for failing to protect Plaintiffs from third-party users on the site. For that reason, courts have consistently held that such claims are barred by [CDA] 230. This has been the unanimous view of other courts confronted with claims alleging that defectively designed internet products allowed for transmission of harmful third-party communications.").

knowingly facilitate human trafficking on its platforms.[194] Three cases were filed in Texas state court by victims of sex trafficking that, similar to *Force*, were argued to have occurred because Facebook knowingly aided sex traffickers by connecting them with and permitting the grooming of minor girls on the company's platforms.[195] The court held that Facebook "affirmatively acted" in the promotion of the human trafficking beyond the scope of CDA 230, writing: "We find it highly unlikely that Congress, by prohibiting treatment of internet companies 'as . . . publisher[s],' sought to immunize those companies from *all* liability for the way they run their platforms, even liability for their own knowing or intentional acts as opposed to those of their users."[196] Finding Facebook's recommendation algorithm an affirmative action by the platform, the Texas court determined the plaintiffs' state claim under a sex trafficking statute could proceed.[197]

Just four months after the Supreme Court of Texas issued its decision at odds with the Second and Ninth Circuits' interpretation of CDA 230 applicability to algorithmic action, the Ninth Circuit issued another decision solidifying the view among federal circuits that platforms are immunized from liability for their algorithmic action. In *Gonzalez v. Google LLC*, family members of victims of terrorist attacks in Paris, Istanbul, and California sued Google for permitting ISIS members to upload content on Google-owned YouTube, which the platform then recommended to users, resulting in their radicalization to commit the fatal acts.[198] The *Gonzalez* court approvingly cited *Force* and held that, as long as YouTube recommended user-developed content equally—that is, treated harmful user-developed content the same as other user-developed content—the platform's recommendation algorithm was a "neutral tool" immune under CDA 230.[199] The victims' families appealed and, in October 2022,

194. *Id.* at 101 ("[CDA] 230(c) does not bar Plaintiffs' claims alleging Facebook's affirmative acts in violation of [the state statute].").

195. *Id.* at 83.

196. *Id.* at 98 (alteration in original).

197. *Id.* at 101. The Supreme Court of Texas also considered the impact of a recent CDA 230 amendment specifically clarifying that CDA 230 immunity does not apply to online facilitation of sex trafficking as supplementary justification for its reading of CDA 230. *Id.* at 99–101 (analyzing the impact of FOSTA on CDA 230 immunity).

198. 2 F.4th 871, 880–85 (9th Cir. 2021).

199. *Id.* at 896 ("[A] website's use of content-neutral algorithms, without more, does not expose it to liability for content posted by a third-party.").

the Supreme Court of the United States granted certiorari in *Gonzalez*.[200]

For the first time since its 1996 enactment, CDA 230 will be interpreted by the Supreme Court at a pivotal moment—a moment of profound offline impacts of online speech.[201] The use of algorithms and AI to moderate and amplify content, as well as to connect users across platforms, put up against a robust CDA 230 immunity shield, has complicated the judiciary's attempt to address the offline impacts of online speech. Supreme Court guidance is overdue.

B. JUDICIAL INTERPRETATION OF CDA 230 HAS OVEREXTENDED THE LAW

CDA 230 immunity has been interpreted to extend beyond its original purposes and courts have contorted the law into what is essentially a get-out-of-jail-free card for online activity beyond speech—including algorithmic action.[202] Additionally, most cases pursued against platforms are dismissed at the pleading stage because of CDA 230's robust shield, halting litigation before discovery can commence and platform conduct can be investigated. In the words of Professor Mary Anne Franks, "[t]he law of cyberspace . . . has unmade the law of real space."[203] Take, for example, the Supreme Court of Wisconsin's extension of CDA 230's shield to an online firearms marketplace in *Daniel v. Armslist, LLC*.[204] In *Armslist*, an e-commerce platform was sued for allowing users to sell firearms to other users without providing

200. *See* SCOTUSBLOG, *supra* note 118 (following the status of *Gonzalez* on appeal).

201. *See supra* Part I.A.3 (discussing the offline impact of online speech).

202. *See* Citron & Wittes, *supra* note 104, at 413–14 (taking stock of various providers and users whose activities have been, or likely will be, immunized from liability under CDA 230).

203. Franks, *supra* note 165, at 14.

204. *See* 926 N.W.2d 710 (Wis. 2019), *cert. denied*, 140 S. Ct. 562 (2019). Radcliffe Haughton was banned from purchasing a firearm because of a restraining order brought by his estranged wife. *Id.* at 714–15. To obtain a firearm, Haughton turned to Armslist.com, a platform connecting unlicensed firearm sellers with buyers, including users who were unable to pass the requisite background check for a legal firearm purchase. *Id.* at 715. Haughton killed four people with the firearm. *Id.* at 714. Yasmeen Daniel, who was inside the building and a witness to the massacre, sued Armslist, asserting claims including negligence, aiding and abetting tortious conduct, public nuisance, and wrongful death. *Id.* at 716.

a background check.[205] The court held, however, that Armslist.com was protected by CDA 230, and that holding Armslist.com liable for facilitating an illegal firearm sale would require holding the platform liable as the "speaker" or "publisher" of user-developed information on the site.[206] In effect, CDA 230 was interpreted in *Armslist* to immunize sellers from liability for facilitating illegal gun sales.[207] Had the transactions taken place offline, no such protection would have been available and a crime would likely be prosecuted.

In another case, CDA 230 proved resilient against products liability claims against the dating app Grindr. In *Herrick v. Grindr, LLC*,[208] a Grindr user alleged the dating app was deceptive about its business practices.[209] Even though the claims were related to Grindr's failure to respond to user requests for the removal of Herrick's personal information from fake dating profiles, each court Herrick petitioned dismissed his claims citing the robust immunity that CDA 230 provided Grindr.[210] Had similar products liability claims been pursued against a company offline, without the forceful CDA 230 shield, a court likely would have found a justiciable issue. In *Herrick*, however, the case was

---

205. *Id.* at 715.

206. *See id.* at 722 ("In this case, Armslist did not develop the content of [the] firearm advertisement, so Armslist is not an information content provider with respect to the advertisement."); *id.* at 726 ("Regardless of Armslist's knowledge or intent, the relevant question is whether Daniel's claim necessarily requires Armslist to be treated as the publisher or speaker of third-party content. Because it does, the negligence claim must be dismissed.").

207. Franks, *supra* note 165, at 14 (citing *Armslist* to show courts have interpreted Section 230 to protect online sites).

208. 306 F. Supp. 3d 579 (S.D.N.Y. 2018). Matthew Herrick was the victim of a dangerous scheme to defraud him on Grindr, a dating app. Herrick's ex-boyfriend had created fake dating profiles of him on the app, posted explicit images, his personal information including his address, phone number, and place of work, and solicited responses from approximately 1,100 men to proposition Herrick for sex over a six-month period. *Id.* at 585. Despite contacting Grindr approximately 100 times for help removing the user-developed content maliciously posted without his consent, he received no response. *Id.* In his complaint, Herrick did not allege that Grindr be held liable for the substantive message of the user-developed content, but rather for its failure to help end the violent situation. *Id.* at 585–86.

209. Carrie Goldberg, Herrick v. Grindr*: Why Section 230 of the Communications Decency Act Must Be Fixed*, LAWFARE (Aug. 14, 2019), https://www.lawfareblog.com/herrick-v-grindr-why-section-230-communications-decency-act-must-be-fixed [https://perma.cc/6WY2-U6ZA] (describing the *Herrick* case from the perspective of the lawyers who prosecuted it).

210. *Id.*

dismissed—with CDA 230 raised as an affirmative defense without any showing that the elements of the defense had been met.[211] *Armslist* and *Herrick* are illustrative of a major issue with judicial interpretations of CDA 230—its extension of immunity to broadly encompass platforms' conduct and courts' brazen dismissal of varied, nuanced cases before considering the merits of the claims.

CDA 230 was written from the traditionally dyadic Old School Speech Regulation framework—to regulate speech, target the speaker or publisher of the message. The reality of the modern internet is far more complex, attributable, at least in part, to the proliferation of algorithms and AI to amplify speech, recommend content, and moderate content. CDA 230 was not enacted by the legislature, nor has it been interpreted by most courts, with an understanding that platforms like Facebook and Twitter act and influence the online marketplace of ideas far beyond classic editorial decisions.[212] Therefore, CDA 230 immunizes platform activity beyond the activity it was originally intended to protect.

It is likely that anyone aware of CDA 230 has an opinion about CDA 230. It is a divisive statute because the consequences are so large. There are meritorious arguments supporting CDA 230 as enacted which are discussed below. Ultimately, however, this Note has demonstrated too many important reasons for its amendment.

---

211. *Id.* ("Usually, to benefit from an affirmative defense like Section 230, a defendant has the burden of proving it satisfies the elements of that defense. Grindr would have needed to serve an answer claiming it was immune under Section 230 and allege all three of the statute elements for the company to get the enormous benefit of immunity—that it was (1) 'an interactive computer service' (2) being 'treated as a publisher' of (3) 'information provided by another information content provider.' Instead, contrary to procedural rules but nevertheless common in Section 230 cases, the judge saved Grindr that step by dismissing the case before Grindr had filed a single pleading.").

212. Citron & Wittes, *supra* note 104, at 413 ("Companies have too limited an incentive to insist on lawful conduct on their services beyond the narrow scope of their terms of service. They have no duty of care to respond to users or larger societal goals. They have no accountability for destructive users of their services, even when they encourage those uses.").

IV. TO REFORM OR NOT TO REFORM CDA 230? THAT IS THE QUESTION

Since "creating" the internet, CDA 230 has fallen from the public's graces. Even in the contentious and polarized 2020 election season, both presidential nominees agreed on one issue—the repeal of CDA 230.[213] This Part introduces scholarly perspectives on CDA 230 and proposed solutions to move towards a healthier online marketplace of ideas. Some scholars, such as Professor Eric Goldman, argue that CDA 230 is fine as is and ought to be left alone.[214] Many scholars propose solutions, less drastic than complete repeal, to the internet's problems by narrowing and focusing the statute's interpretation. These solutions include an interpretative shift of the statute to exclude "Bad Samaritans" from immunity and amending CDA 230's statutory language to tailor its immunity to apply exclusively to online speech, as well as modernizing the statute to account for algorithmic action across platforms.

A. THE CASE FOR LEAVING CDA 230 ALONE

Despite the offline harms of online speech, some CDA 230 advocates warn against changing the law. Professor Eric Goldman credits the statute with four major benefits: (1) job creation in the U.S. internet and technology sector and a corresponding U.S. GDP increase,[215] (2) the promotion of small businesses by deterring frivolous legal action brought against them, (3) strengthening the market by creating online marketplaces and improving consumer decision-making with the emergence of consumer reviews on websites like Yelp, and (4) fostering free

---

213. *See* Editorial Bd., *Joe Biden*, N.Y. TIMES (Jan. 17, 2020), https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html [https://perma.cc/SP6V-EGF7] (quoting then-presidential candidate Joe Biden, "Section 230 should be revoked, immediately should be revoked . . . ."); Bobby Allyn, *As Trump Targets Twitter's Legal Shield, Experts Have a Warning*, NPR (May 30, 2020), https://www.npr.org/2020/05/30/865813960/as-trump-targets-twitters-legal-shield-experts-have-a-warning [https://perma.cc/ZK5P-7MH9] (citing to former-President Donald Trump's tweet reading "REVOKE 230!").

214. *See, e.g.*, Goldman, *supra* note 148.

215. Christian M. Dippon, *Economic Value of Internet Intermediaries and the Role of Liability Protections*, NERA ECON. CONSULTING 2 (June 5, 2017), https://cdn1.internetassociation.org/wp-content/uploads/2017/06/Economic-Value-of-Internet-Interm [https://perma.cc/V4AH-WP8D] (finding that the weakening of internet safe harbor provisions, such as CDA 230, would eliminate 425,000 jobs in the U.S. internet sector and decrease U.S. gross domestic product by $44 billion annually).

speech for all, including marginalized voices from groups typically excluded from public discourse.[216] Amending CDA 230, Goldman says, would "shrink the internet"[217] and, thus, these enumerated benefits.

Goldman's points are well-taken. Imagining a world without modern internet is an impossible mental exercise, but the economy would undoubtedly shrink, and the job market would lack a sizable sector of opportunity.[218] It is also easy to take for granted the way the internet has transformed online public discourse to include voices of the ignored or silenced, however, these voices belong to the same users who are most likely to experience online harassment and abuse.[219] So, as diversity of participation in public discourse has increased with the internet, the exposure to and experience of harassment and abuse of users from historically marginalized identities and backgrounds has also increased.

Goldman makes three compelling arguments for why amending CDA 230 would cause more harm than good for online speech. First, it is impossible to make all users happy with content moderation practices, and revoking CDA 230 immunity would be "litigation-bait" for users who are simply dissatisfied with a platform's content moderation decisions.[220] It is true that it is impossible to make all internet users happy with internet regulation and content moderation practices, but detailed regulatory efforts are not passed by consensus, and the democratic process often leaves constituents disappointed. A full repeal of CDA 230, without further action, would likely allow "litigation-bait" for individuals either dissatisfied with a platform's content

216. Letter from Eric Goldman & David S. Levine, Professors, to Members of Cong. (Mar. 9, 2020), https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=3164&context=historical [https://perma.cc/S8V2-NL5R].

217. Eric Goldman, *The Plan to Blow Up the Internet, Ostensibly to Protect Kids Online*, CAPITOL WKLY. (Aug. 18, 2022), https://capitolweekly.net/the-plan-to-blow-up-the-internet-ostensibly-to-protect-kids-online [https://perma.cc/A59Q-S8J4].

218. Dippon, *supra* note 215.

219. *See generally* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014) (investigating the significant, personal cyber-attacks happening online and proposing practical ways to prevent and punish online harassment and abuse).

220. Eric Goldman & Jess Miers, *Why Can't Internet Companies Stop Awful Content?*, ARS TECHNICA 3 (Nov. 27, 2019), https://arstechnica.com/tech-policy/2019/11/why-cant-internet-companies-stop-awful-content [https://perma.cc/R3WU-JURD]. See Goldman, *supra* note 148, at 42–44, for more of the procedural benefits of CDA 230.

moderation practices or user-developed content they do not like or are harmed by.[221] This Note, however, does not argue for a repeal of CDA 230 in the absence of a replacement regulatory regime. In fact, the scholarly proposals described below advocate, not for a repeal of CDA 230, but instead a statutory amendment to tailor its scope to be both consistent with congressional intent and in consideration of the realities of the modern internet.[222]

Second, Goldman worries that getting tough on platforms' content moderation would create a barrier to entry for market newcomers, the small tech start-ups that would be unable to afford compliance costs should more methodical regulation replace CDA 230.[223] It is true that CDA 230 has thwarted frivolous lawsuits that could have taken down small entrepreneurial digital ventures.[224] However, the reality of the marketplace for platforms is that there is significant concentration already[225] and reforming CDA 230, as opposed to repealing it without a substitute, can continue to protect small to mid-size technology companies from frivolous lawsuits.

Third, content moderation, Goldman argues, can never eradicate anti-social behavior[226] online because such behavior is part of the human condition.[227] Anti-social behavior will likely occur online no matter the repeal or amendment of CDA 230. The objective of this Note and the proposals below is not to eradicate users' anti-social behavior, but to begin holding platforms

221. *See* Lawpod, *OA650: If Fake Insulin Tweet Tanked Stock, Could Eli Lilly Sue?*, OPENING ARGUMENTS (Nov. 21, 2022), https://openargs.com/oa650-if-fake-insulin-tweet-tanked-stock-could-eli-lilly-sue [https://perma.cc/A9NM-4P7T] (explaining that a complete repeal of CDA 230 would likely lead to more false information online).

222. *See infra* Part IV.B.

223. Goldman & Miers, *supra* note 220.

224. Ashley Johnson & Daniel Castro, *Fact-Checking the Critiques of Section 230: What Are the Real Problems?*, INFO. TECH. & INNOVATION FOUND. (Feb. 22, 2021), https://itif.org/publications/2021/02/22/fact-checking-critiques-section-230-what-are-real-problems [https://perma.cc/LE2T-EY8W].

225. *See, e.g.*, Arwa Madawi, *Zuckerberg Now Runs Not a Business but an Empire. It's Time to Strike Back*, GUARDIAN (Sept. 30, 2018), https://www.theguardian.com/technology/2018/sep/30/facebook-zuckerberg-empire-instagram-whatsapp-antitrust [https://perma.cc/NQ7M-7BSQ].

226. Anti-social behavior is behavior that violates the basic rights of others. *See* Susan D. Calkins & Susan P. Keane, *Developmental Origins of Early Antisocial Behavior*, 21 DEV. & PSYCHOPATHOLOGY 1095, 1095 (2009).

227. Goldman & Miers, *supra* note 220.

responsible when their AI and algorithms amplify, encourage, or exacerbate that anti-social behavior.

Even supporters of CDA 230, however, are puzzled by how algorithms and AI decision making factor into the analysis. For example, Goldman has acknowledged that Facebook's AI has potentially caused racial discrimination in its advertising.[228] Facebook has been subjected to much scrutiny for its potential to algorithmically classify users into racial and ethnic affinity groups, and target housing and employment ads based on that classification.[229] Because the platform and its advertising practices are the cause of the discrimination, not the content of the advertising, Goldman says "it's a tough legal question" to figure out how CDA 230 applies to the situation.[230] Fortunately, scholars have already begun to identify steps towards resolving that difficult legal question.

B. RETHINKING CDA 230

This section summarizes scholarly calls to make crucial modifications to CDA 230's scope and interpretation. These proposals balance the interest in maintaining the internet's safe harbor with the recognition that CDA 230's immunity has been overextended and caused harm inconsistent with other motivating factors—such as moderating obscene content online—behind its enactment.

1. Revoke CDA 230 Immunity for "Bad Samaritans"

CDA 230 was established to both encourage content moderation and shield sites from liability if they failed to perfectly moderate that content.[231] The result has been expansive immunity, reducing incentives for sites to regulate content in good faith.[232] This shield from liability has catalyzed an explosion of platforms aimed at encouraging and facilitating bad behavior

---

228. Eric Goldman, *My Interview with Mathew Ingram Regarding Section 230*, TECH. & MKTG. L. BLOG (June 28, 2020), https://blog.ericgoldman.org/archives/2020/06/my-interview-with-mathew-ingram-regarding-section-230.htm [https://perma.cc/8285-KR74].

229. Jinyan Zang, *Solving the Problem of Racially Discriminatory Advertising on Facebook*, BROOKINGS (Oct. 19, 2021), https://www.brookings.edu/research/solving-the-problem-of-racially-discriminatory-advertising-on-facebook [https://perma.cc/4KH6-VNVL].

230. Goldman, *supra* note 228.

231. *See supra* Part II.

232. Citron & Wittes, *supra* note 104.

online, a purpose far from and inconsistent with CDA 230's origins as a statute aimed to censor "offensive" material and protect against abuse.[233] Professor Citron and Benjamin Wittes of the Brookings Institute propose a practical interpretative shift of CDA 230, consistent with its statutory purpose.[234] Citron and Wittes argue that CDA 230(c)(1)—which extends immunity from liability to "Good Samaritan" platforms that, despite a good faith effort, fail to perfectly moderate content—should be interpreted consistent with its congressional purpose and be limited to platforms that put forth a good faith effort to moderate content.[235] This interpretative shift would exclude "Bad Samaritans"—platforms that facilitate destructive online abuse or know they are principally used for that purpose—from CDA 230 immunity.[236]

This solution is practical and consistent with the original purpose of CDA 230. The key challenge to this proposal is definitional. Which sites qualify as a "Bad Samaritan?"[237] For example, TheDirty is a gossip website for user-posted, well, "dirt."[238] In 2014, the Sixth Circuit ruled that TheDirty was privileged to CDA 230 immunity despite hosting salacious and defamatory posts.[239] Just as courts have had to make similar nebulous distinctions before—for example, distinguishing salacious gossip

---

233. *Id.* at 403.

234. *Id.* at 416.

235. *Id.* This shift is possible through the courts or in a statutory amendment. *Id.* at 418–19. Citron and Wittes propose the following statutory language: "No provider or user of an interactive computer service that *takes reasonable steps to prevent or address unlawful uses of its services* shall be treated as the publisher or speaker of any information provided by another information content provider *in any action arising out of the publication of content provided by that information content provider*." *Id.* at 419 (emphasis in original).

236. *Id.* at 416. Citron's and Wittes's suggestion has not been ignored. In 2020, the Trump Administration's Department of Justice sent legislation to Congress outlining proposed CDA 230 reform. *See* Press Release, U.S. Dep't of Just., The Justice Department Unveils Proposed Section 230 Legislation (Sept. 23, 2020), https://www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation [https://perma.cc/3428-72XT]. Among the revisions was a provision removing CDA 230 liability for bad actors, defined as platforms that knowingly facilitate criminal behavior or knowingly fail to remove content that violates criminal law. *Id.*

237. KOSSEFF, *supra* note 22, at 221 ("How would Congress—or the courts—draw a line that distinguishes TheDirty from Yelp? Would that line be clear enough to provide certainty to companies that are building new businesses that are based on third-party content?").

238. *Id.* at 220.

239. Jones v. Dirty World Ent. Recordings, 755 F.3d 398 (6th Cir. 2014).

from newsworthy subject matter in traditional media[240]—courts can do so with "Bad Samaritans." The definition will likely be imperfect, but isn't much of the law? It is likely that, if the law is limited to exclude "Bad Samaritans," best practices and principles will emerge from courts in time.[241]

This definitional question is further complicated when considering AI and its impact on platforms that are not inherently "Bad Samaritans." Facebook is not primarily used for racially discriminatory purposes, but its technology can be used that way.[242] Additionally, the platform intentionally acted to perpetuate division between users on its platform for ad revenue.[243] It is unclear how Facebook and similar platforms that are not primarily used for nefarious and harmful deeds would fit within the "Bad Samaritan" definition.

Citron and Wittes further propose a "reasonable standard of care" for platforms.[244] Citron further developed this standard in partnership with Franks, advocating for an elastic standard that permits platforms' flexibility to develop norms for content moderation practices.[245] Just as CDA 230 was originally written with imperfect content moderation in mind, a reasonable standard of care would consider the same. All that would be required under this standard is for a platform to take reasonable steps to address unlawful and abusive user-posted speech.[246] A glaring pitfall of this standard is that the line between, for example, incitement of violence and constitutionally protected free speech, is tenuous and imprecise. However, Courts have had to make similar determinations in specialized fields before and are equipped to interrogate the reasonableness of a platform's policies.[247]

Interpreting CDA 230 to exclude "Bad Samaritans" from its immunity is a reasonable proposed solution to the issue of online

240. Sarah Hinchliff Pearson, *Who Decides What Is Newsworthy? Journalists vs. the Legal System*, CTR. FOR INTERNET & SOC'Y (June 9, 2009), http://cyberlaw.stanford.edu/blog/2009/06/who-decides-what-newsworthy-journalists-vs-legal-system [https://perma.cc/UXW2-YV4P].

241. Citron & Franks, *supra* note 88, at 71–73.

242. *See supra* notes 228–30 and accompanying text.

243. *The Facebook Files*, *supra* note 94 (describing Facebook documents laying out the company's decision to implement a new algorithm in 2018 that increased user engagement by increasing sensational, divisive content and misinformation).

244. Citron & Wittes, *supra* note 104, at 423.

245. Citron & Franks, *supra* note 88, at 74.

246. *Id.*

247. *Id.* at 72.

speech regulation. There is a definitional issue with the proposal, however. Any definition of "Bad Samaritan" should include platforms that are primarily used for unlawful or abusive conduct online, as well as platforms that knowingly encourage unlawful or abusive online conduct through AI or algorithmic action. Enforcing a "reasonable standard of care" for platforms is a strong solution pulling from an established common law legal standard. However, in reality, it would place a heavy and complicated burden on courts to both grasp complex technological concepts and draw lines in some of the most profound and unclear questions in First Amendment law.

2. Replace "Information" with "Speech" in the Statutory Language to Narrow the Immunity Shield

A second interpretative shift could clarify CDA 230 to apply only to speech, not all online behavior.[248] Courts have broadened CDA 230 immunity to online behavior beyond speech.[249] Citron and Franks propose a minor amendment to the language of CDA 230 that would have a substantial impact on the statute's scope.[250] By replacing "information" with "speech" in the statutory text, courts would no longer be able to short circuit the speech inquiry.[251] Alternatively, Citron and Franks propose that courts determine, first, if the facts underlying the claims arise from user-developed content and, if they do, whether it is constitutionally protected speech.[252] This inquiry would slow down the court's deliberation when faced with a CDA 230 affirmative defense in a case dismissal.

A drawback of this reform is that, even if the scope of CDA 230 is limited to instances of speech at the exclusion of other online behavior, there remains no regulatory mechanisms for platforms' conduct, such as its advertising strategies.[253] Without protection under CDA 230, however, platforms would likely be liable under applicable laws as offline companies are. Unless a platform's AI or algorithmic action is considered constitutionally

---

248. Citron & Franks, *supra* note 88.

249. *See supra* Part III.

250. Citron & Franks, *supra* note 88, at 69–70.

251. *Id.*

252. *Id.* at 45.

253. *See supra* notes 228–30 and accompanying text.

protected free speech,[254] this type of conduct will fall outside the bounds of CDA 230 immunity.

These proposals for interpretative shifts of CDA 230 immunity are compelling because they center the importance of enabling internet innovation and balance it against the wellbeing and safety of internet users, and the public at large. These proposals encourage returning to CDA 230's original purpose and narrowing its scope consistent with its enactment.

3. Adding a Flexible Modern Legal Standard to Address Algorithmic Influence

One deficiency of CDA 230 is its application of the dyadic Old School Speech Regulation framework for a far more dynamic communication medium.[255] Platforms are not "speakers" or "publishers." Platforms have capabilities that far exceed the editorial functions of those traditional roles, such as the algorithmic action made by the platforms to recommend or amplify certain user-developed content.[256] An effective modernization of CDA 230, or any internet law that proceeds CDA 230, should acknowledge those capabilities. One way to accomplish this is to develop a new legal standard to account for the unique roles and capabilities of platforms—Algorithm-Based Republisher (ABR).[257] Though platforms do not post the content they host, they do, in many ways, determine what internet users see online through amplification and recommendations. This is a capability and impact beyond that of a common publisher of information.[258]

254. There is debate about whether computer code is protected speech under the First Amendment. *See* Claire Hodges, *Decoding Speech: The Implications of Giving Computer Code First Amendment Protections*, BROWN POL. REV. (Mar. 7, 2022), https://brownpoliticalreview.org/2022/03/decoding-speech [https://perma.cc/2T63-K8VB] (describing the Ninth Circuit's ruling that computer code is protected speech and implications should the Supreme Court adopt similar logic).

255. *See supra* Part II.

256. Grafanaki, *supra* note 39, at 126 ("Big Data technologies now allow for precise tracking and analysis at the moment of content consumption. They also allow for instantaneous adjustment of the content selection based on the feedback. Traditional *editors* never had that kind of power.") (emphasis in original).

257. Seema Ghatnekar, Note, *Injury by Algorithm*, 33 LOY. L.A. ENT. L. REV. 171, 201 (2013).

258. Michael P. Bennett & Ryan T. Sulkin, *Ninth Circuit Tightens the Belt on Immunities for Online Publishers of User-Generated Content*, LEXOLOGY (June 8, 2007), https://www.lexology.com/library/detail.aspx?g=9a998c44-1ab3-4124-9d3d-4ae2f1a5dbec [https://perma.cc/M2HV-YPLA].

An ABR would be situated between a class of sites that authors and publishes its own content, such as the New York Times website, and a site that hosts user-developed content as a neutral tool, without any impact on how that content is spread or received, such as a digital bulletin board like Craigslist. An ABR's liability, therefore, would fall between the total liability or total immunity currently available under CDA 230.[259] Presently, such a concept is not recognized in U.S. law, so platforms that act as ABRs are immunized from liability under CDA 230.[260] A few countries, however, have attributed liability to Google when its algorithm directed users to suggestive and harmful content.[261] It is difficult to draw a bright line distinction between publishing and the amplification of ABR,[262] but courts are accustomed to handling complexity and can, over time, determine what liability an ABR has for elevating harmful content.

Developing a unique legal category for ABRs seems reasonable enough, however there are a couple obvious pitfalls. First, it is not clear what metrics a court could use to determine an ABR's liability. Perhaps the court could engage in a hypothetical exercise of how impactful certain content would be had an algorithm not catalyzed its spread, but that seems too speculative for fair and consistent outcomes. Second, technology is ever evolving, and quickly. Introducing a new legal standard in this context might prove completely ineffective in a couple years' time when new functionality and mechanics develop to support online speech.

CDA 230 has enabled an incredible internet and, with it, a stronger economy and inclusion of marginalized groups in public discourse. The CDA 230 reform proposals in this Part, however, offer practical suggestions for how to enhance the internet by regulating it to encourage and facilitate robust and respectful online speech.

CONCLUSION

This Note has argued that there is a vacuum in the legal landscape pertaining to online speech regulation. The employ-

---

259. Ghatnekar, *supra* note 257, at 202.

260. *Id.*

261. *Id.* (citing cases in Australia, Japan, and France where Google has been held liable for its algorithm).

262. Keller, *supra* note 67, at 233.

ment of algorithms and AI to the digital infrastructure have altered the way platforms interact with user-developed content and, therefore, the legal analysis for liability should be altered. This Note has explained the modern internet's characteristics that make online speech regulation so difficult, the challenge courts have in interpreting CDA 230 in the face of algorithmic harms, and analyzed three possible solutions to reform CDA 230.

The internet is not hopeless. The internet is flawed and harmful and full of opportunity and connection. CDA 230 has been instrumental to creating the internet of today—the good, the bad, the ugly. As the internet innovates and evolves, however, so should the law. This Note has surveyed scholarly solutions to amending CDA 230, and now it is time for the government to act. A new or reformed CDA 230 to regulate online speech should hold "Bad Samaritans" responsible for explicitly or knowingly perpetuating harmful content on their platforms; be narrow enough in scope to regulate online speech, rather than all activity online; and add a flexible modern legal standard so that courts may more proficiently and consistently assess algorithmic harm on platforms. The internet has introduced new challenges and expanded the ability to connect and build community. The law must meet those challenges and keep that community safe.