Fordham Intellectual Property, Media and Entertainment Law Journal

Volume 32 *XXXII*
Number 1

Article 2

2021

Targeting Exceptions

Michal Lavi

Recommended Citation

Michal Lavi, *Targeting Exceptions*, 32 Fordham Intell. Prop. Media & Ent. L.J. 65 (2021).
Available at: https://ir.lawnet.fordham.edu/iplj/vol32/iss1/2

Targeting Exceptions

Cover Page Footnote

* Michal Lavi Ph.D (Law); Research Fellow at the Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crises and Technology. I thank Emily Cooper. Special thanks are due to Daniel Levin, Laura Rann, Caroline Vermillion, and their colleagues on the Fordham Intellectual Property, Media & Entertainment Law Journal staff for their helpful comments, suggestions, and outstanding editorial work that profoundly improved the quality of this Article. I dedicate this Article to the memory of my mother, Aviva Lavi, who died suddenly and unexpectedly. My mother taught me to love knowledge and gave me the strength to pursue it. She will always be loved, remembered, and dearly missed.

Targeting Exceptions

Michal Lavi[*]

On May 26, 2020, the forty-fifth President of the United States, Donald Trump, tweeted: "There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed." Later that same day, Twitter appended an addendum to the President's tweets so viewers could "get the facts" about California's mail-in ballot plans and provided a link. In contrast, Facebook's CEO Mark Zuckerberg refused to take action on President Trump's posts. Only when it came to Trump's support of the Capitol riot did both Facebook and Twitter suspend his account. Differences in attitude between platforms are reflected in their policies toward political advertisements. While Twitter bans such ads, Facebook generally neither bans nor fact-checks them.

The dissemination of fake news increases the likelihood of users believing it and passing it on, consequently causing tremendous reputational harm to public representatives, impairing the general public interest, and eroding long-term democracy. Such dissemination depends on online intermediaries that operate platforms, facilitate dissemination, and govern the flow of information by moderating, providing algorithmic recommendations, and targeting third-party

[*] Michal Lavi Ph.D (Law); Research Fellow at the Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crises and Technology. I thank Emily Cooper. Special thanks are due to Daniel Levin, Laura Rann, Caroline Vermillion, and their colleagues on the *Fordham Intellectual Property, Media & Entertainment Law Journal* staff for their helpful comments, suggestions, and outstanding editorial work that profoundly improved the quality of this Article.

I dedicate this Article to the memory of my mother, Aviva Lavi, who died suddenly and unexpectedly. My mother taught me to love knowledge and gave me the strength to pursue it. She will always be loved, remembered, and dearly missed.

advertisers. Should intermediaries bear liability for moderating or failing to moderate? And what about providing algorithmic recommendations and allowing data-driven advertisements directed toward susceptible users?

In A Declaration of the Independence of Cyberspace, John Perry Barlow introduced the concept of internet exceptionalism, differentiating it from other existing media. Internet exceptionalism is at the heart of Section 230 of the Communications Decency Act, which provides intermediaries immunity from civil liability for content created by other content providers. Intermediaries like Facebook and Twitter are thereby immune from liability for content created by users and advertisers. However, Section 230 is currently under attack. In 2020, Trump issued an "Executive Order on Preventing Online Censorship" that aimed to limit platforms protections against liability for intermediary-moderated content. Legislative bills seeking to narrow Section 230's scope soon followed. From another direction, attacks on the overall immunity provided by Section 230 emerged alongside the transition from an internet society to a data-driven algorithmic society—one that changed intermediaries' scope and role in information dissemination. The changes in the utility of intermediaries requires reevaluation of their duties; that is where this Article steps in.

This Article focuses on dissemination of fake news stories as a test case. It maps the roles intermediaries play in the dissemination of fake news by hosting and moderating content, deploying algorithmically personalized recommendations, and using data-driven targeted advertising. The first step toward developing a legal policy for intermediary liability is identifying the different roles intermediaries play in the dissemination of fake news stories. After mapping these roles, this Article examines intermediary liability case law and reflects on internet exceptionalism's current approach and recent developments. It further examines normative free speech considerations regarding intermediary liability within the context of the different roles they play in fake news dissemination and argues that the liability regime must correspond with the intermediary's role in dissemination. By targeting exceptions to internet exceptionalism, this Article outlines a nuanced framework for intermediary liability. Finally, it proposes subjecting intermediaries to transparency *obligations regarding moderation practices and imposing duties to conduct algorithmic impact assessments as part of consumer protection regulation.*

INTRODUCTION

On May 26, 2020, the forty-fifth President of the United States, Donald Trump, tweeted "The Governor of California is sending Ballots to millions of people, anyone . . . living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!"[1] Later that same day, Twitter appended an addendum to the former President's tweets that stated viewers could "get the facts" about California's

[1] Trump's tweet is no longer available due to his suspension. For a journal article citing this Tweet, see Todd Spangler, *Twitter Adds Warning Label to Donald Trump's False Tweets for the First Time*, VARIETY (May 26, 2020, 3:22 PM), https://variety.com/2020/digital/news/twitter-adds-warning-label-donald-trumps-false-tweets-for-first-time-1234616642/ [https://perma.cc/N43M-ULL7].

mail-in ballot plans and provided a link to the information.[2] Twitter continued flagging Trump's tweets during the 2020 election cycle and even continued after the election concluded.[3]

In contrast to Twitter's moderation practices, Facebook avoided labeling Trump's post.[4] Despite Facebook employees' protests regarding the company's lack of response to Trump's posts,[5] Facebook CEO Mark Zuckerberg continued to defend his decision not to interfere with the President's posts.[6] However, after market pressure from advertisers,[7] Facebook announced it would flag all "newsworthy" posts by politicians that violate its community rules.[8]

[2] *See Politics: Trump Makes Unsubstantiated Claim That Mail-In Ballots Will Lead to Voter Fraud*, TWITTER (May 26, 2020), https://twitter.com/i/events/1265330601034256384 [https://perma.cc/5RHN-Y8G9].

[3] *See* Kim Lyons, *Twitter Flags President Trump's Tweets About Ballot-Counting*, VERGE (Nov. 7, 2020, 10:19 AM), https://www.theverge.com/2020/11/7/21554013/twitter-flags-president-trumps-tweets-votes-counted-election-pennsylvania [https://perma.cc/4EWK-KKK6]; *see also Trump Falsely Claims Victory on Twitter Just Ahead of Biden Win*, QUINT (Nov. 7, 2020, 10:24 PM), www.thequint.com/news/world/won-by-a-lot-president-trump-falsely-declares-victory-on-twitter-again [https://perma.cc/3LXV-YKLS] (referring to Trump's misleading tweet "I WON THIS ELECTION, BY A LOT!").

[4] Audrey Conklin, *Facebook Won't Label Trump 'Mail-In Ballot' Post Like Twitter*, FOX BUS. (May 27, 2020), https://www.foxbusiness.com/technology/facebook-twitter-label-trump-ballot [https://perma.cc/Q3HB-A33M].

[5] Fanny Potkin et al., *Facebook Staffers Walk Out Saying Trump's Posts Should be Reined In*, REUTERS (June 1, 2020, 07:51 AM), https://www.reuters.com/article/us-facebook-trump-employee-criticism/facebook-staffers-walk-out-saying-trumps-posts-should-be-reined-in-idUSKBN2382D0 [https://perma.cc/YHJ4-E8RF].

[6] Ashutosh Bhagwat, *The Law of Facebook*, 54 U.C. DAVIS L. REV. 2353, 2363–64 (2021) ("Facebook decided *not* to restrict targeted political advertising or to fact-check political ads (as opposed to commercial ads) . . . ."); *see also* Tony Romm et al., *Facebook Won't Limit Political Ad Targeting or Stop False Claims Under New Ad Rules*, WASH. POST (Jan. 9, 2020), https://www.washingtonpost.com/technology/2020/01/09/facebook-wont-limit-political-ad-targeting-or-stop-pols-lying/ [https://perma.cc/K5VM-AEBN]; Elizabeth Dwoskin, *Mark Zuckerberg Defends Decisions on Trump as Facebook Employee Unrest Grows*, WASH. POST (June 2, 2020), https://www.washingtonpost.com/technology/2020/06/02/facebook-zuckerberg-trump-defense/ [https://perma.cc/B3F3-CJWK].

[7] Tiffany Hsu & Gillian Friedman, *CVS, Dunkin', Lego: The Brands Pulling Ads from Facebook Over Hate Speech*, N.Y. TIMES, https://www.nytimes.com/2020/06/26/business/media/Facebook-advertising-boycott.html [https://perma.cc/Z34B-QJLJ] (July 7, 2020).

[8] Barbara Ortutay, *Facebook to Label All Rule-Breaking Posts—Even Trump's*, AP NEWS (June 26, 2020), https://apnews.com/article/donald-trump-us-news-ap-top-news-mark-zuckerberg-ca-state-wire-b38818f48561889452c77fe736646454 [https://perma.cc/RPA4-QHLK].

The differences between Facebook's and Twitter's treatment of third-party content were again made apparent in a similar circumstance during the election. The Trump campaign released a thirty-second video advertisement accusing opponent Joe Biden of allegedly promising to pay Ukraine to fire a prosecutor who investigated a company with ties to Biden's son, Hunter Biden.[9] CNN refused to air the advertisement, finding no evidence supporting the claims.[10] Facebook, however, allowed the advertisement to remain on the platform and declined the Biden campaign's request to remove it.[11] Thus, Facebook allowed this fake story to spread widely and proliferate.[12]

However, on January 5 and 6 of 2021, Trump used social media to encourage a riot at the United States Capitol that was planned by his supporters in an effort to overturn the 2020 presidential election results by calling his supporters to "be there and to be wild."[13] As a consequence, both Facebook and Twitter barred Trump's social media accounts.[14]

[9] *See* Eugene Kiely & Robert Farley, *Fact: Trump TV Ad Misleads on Biden and Ukraine*, FACTCHECK.ORG (Oct. 9, 2019), https://www.factcheck.org/2019/10/fact-trump-ad-misleads-on-biden-and-ukraine/ [https://perma.cc/9R3D-FBKV].

[10] Stephanie Baker et al., *On Bidens and Ukraine, Wild Claims with Little Basis*, BLOOMBERG (Oct. 9, 2019, 9:20 AM), https://www.bloomberg.com/news/articles/2019-10-09/on-bidens-and-ukraine-wild-claims-with-little-basis-quicktake [https://perma.cc/Q5JZ-XVWP].

[11] *See* Emily Stewart, *Facebook Is Refusing to Take Down a Trump Ad Making False Claims About Joe Biden*, VOX (Oct. 9, 2019, 2:30 PM), https://www.vox.com/policy-and-politics/2019/10/9/20906612/trump-campaign-ad-joe-biden-ukraine-facebook [https://perma.cc/J7ZK-9JMQ].

[12] *Id. But see, e.g.*, Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, 1 J. FREE SPEECH L. 71, 92 (2021) ("Facebook's case is instructive for how to think about the problem. Facebook argues that it does not want to be the arbiter of public discourse. In fact, it already is the arbiter of public discourse worldwide . . . .Facebook well understands this: [i]t takes down lies about election dates and polling places . . . .").

[13] Dan Barry & Sheera Frenkel, *'Be There. Will Be Wild!': Trump All but Circled the Date*, N.Y. TIMES, https://www.nytimes.com/2021/01/06/us/politics/capitol-mob-trump-supporters.html [https://perma.cc/4LXY-C6Z6] (July 27, 2021).

[14] Mike Isaac & Kate Conger, *Facebook Bars Trump Through End of His Term*, N.Y. TIMES, https://www.nytimes.com/2021/01/07/technology/facebook-trump-ban.html [https://perma.cc/TNG3-9LQU] (May 18, 2021); *see* Brian Fung, *Twitter Bans President Trump Permanently*, CNN BUS., https://edition.cnn.com/2021/01/08/tech/trump-twitter-ban/index.html [https://perma.cc/DK67-KW84] (Jan. 9, 2021, 2:19 PM).

Candidates in election campaigns, along with their proponents and other stakeholders, disseminate content and fund online political advertisements to find voters to convert into donors, recruit volunteers, and mobilize individuals to vote on Election Day.[15] Fake news stories can be found in both organic content and paid advertisements[16] that gain influence through popularization on online platforms such as Facebook, Twitter, YouTube, and even Google's search engine.

Fake news is amazingly powerful and dangerous when it spreads. Further, it is difficult to clean up the tracks it leaves behind. It has severe consequences on the reputations of public representatives and infringes public interest at large.[17] Fake news pollutes the flow of information as it spills into the digital ecosystem, caused by users spreading such stories widely and extensively.[18] Studies show that fake stories circulate "significantly farther, faster, deeper, and more broadly than the truth"[19] because they hold the audience's attention by eliciting surprise. The more frequently people are exposed to a fake news story, the more credibility ascribed to it.[20] The information begins to seem so true that readers deny its falsity

[15] *See* Daniel Kreiss & Matt Perault, *Four Ways to Fix Social Media's Political Ads Problem— Without Banning Them*, N.Y. TIMES (Nov. 16, 2019), https://www.nytimes.com/2019/11/16/opinion/twitter-facebook-political-ads.html [https://perma.cc/S9LZ-9FX7].

[16] *See, e.g.*, Brian Fung, *Trump Campaign Runs Hundreds of Misleading Facebook Ads Warning of Super Bowl Censorship*, CNN BUS., https://edition.cnn.com/2020/01/24/media/trump-super-bowl-facebook-ad/index.html [https://perma.cc/348N-Y4HR] (Jan. 24, 2020, 8:33 PM).

[17] Cass R. Sunstein*, Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 387, 388 (2020) ("Some falsehoods are harmful. They ruin lives. They lead people to take unnecessary risks or fail to protect themselves against serious dangers.").

[18] *See* Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 112–13 (2019) (using a metaphor comparing "fake news" to "data pollution" that disrupts social institutions and public interests, in a similar manner to environmental pollution).

[19] Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCI. 1146, 1146 (2018), https://www.science.org/doi/abs/10.1126/science.aap9559 [https://perma.cc/9QDD-ETX2]; SINAN ARAL, THE HYPE MACHINE: HOW SOCIAL MEDIA DISRUPTS OUR ELECTIONS, OUR ECONOMY, AND OUR HEALTH—AND HOW WE MUST ADAPT 28 (2020).

[20] Gerd Gigerenzer, *External Validity of Laboratory Experiments: The Frequency-Validity Relationship*, 97 AM. J. PSYCH. 185, 185 (1984) (describing that repeating information creates an illusion of truth).

despite being shown contrary evidence.[21] Further, fake news stories that circulate on social media confirm existing user biases and start a social dynamic of dissemination. Fake news can spread like wildfire when it reaches central "influential entities" in the social network that have more social connections and force than the average user; [22] these influential entities then pass the stories along and increase support for a political candidate, consequently influencing democracy.[23]

Network structures and dynamics within social networks influence how information spreads. Yet, intermediaries that operate social network platforms have an equally influential role in disseminating information. Sacha Baron Cohen recently coined social media platforms the "greatest propaganda machine in history," and this stands to reason.[24]

Intermediaries are more than mere middlemen. They moderate users' content, in turn influencing what users view,[25] value, and

[21] Whitney Phillips, *The Toxins We Carry*, COLUM. JOURNALISM REV. (2019), https://www.cjr.org/special_report/truth-pollution-disinformation.php [https://perma.cc/VY6N-YAWX] ("It shows that when people are repeatedly exposed to false statements, those statements start to feel true, even when they are countered with evidence. In short, a fact check is no match for a repeated lie.").

[22] YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS 284–86 (2018) (explaining that in the 2016 U.S. election campaign, ideological rightwing political news sites, such as Breitbart, adopted fake news and were in fact a springboard for its widespread dissemination. Even though Breitbart is a website, not a social media platform, this example demonstrates the importance the social network's structure. A receptive (ideological) node on a social network can make a difference and explain why negative fake news about Hillary Clinton spread widely, while negative fake news about Donald Trump was disseminated far less.).

[23] *Id.*; Katherine Haenschen & Jay Jennings, *Mobilizing Millennial Voters with Targeted Internet Advertisements: A Field Experiment*, 36 POL. COMMC'N 357, 357–67 (2019) (demonstrating that the internet can be used to access younger people via cookie targeting, reach them with individually targeted advertisements from a local organization, and engage them to vote locally).

[24] *See* Sacha Baron Cohen, *Read Sacha Baron Cohen's Scathing Attack on Facebook in Full: 'Greatest Propaganda Machine in History'*, GUARDIAN (Nov. 22, 2019, 1:10 PM), https://www.theguardian.com/technology/2019/nov/22/sacha-baron-cohen-facebook-propaganda [https://perma.cc/G4FT-2SGZ].

[25] *See, e.g.*, Alex Hern, *Twitter Hides Donald Trump Tweet for 'Glorifying Violence'*, GUARDIAN (May 29, 2020, 12:57 PM), https://www.theguardian.com/technology/2020/may/29/twitter-hides-donald-trump-tweet-glorifying-violence [https://perma.cc/6XZN-8BA5].

repost.[26] Intermediaries have different terms of services and community guidelines.[27] Further, there exists diversity among platforms regarding attitudes toward moderating content. In addition to moderation, intermediaries seek to hold users' attention as long as possible.[28] To do so, they utilize algorithms which personalize and recommend organic content, exposing users to false information and extremist political views.[29] The algorithm learns users' preferences and encourages users to connect with like-minded people, thereby creating "echo chambers" that confirm previously held beliefs.[30] These dynamics not only affect the individual but change the entire social dynamic within a network. [31]

Moreover, intermediaries target political advertisements for profit. Whereas personalizing organic content generally aims to enhance users' engagement by exposing them to relevant content, data-driven targeted advertising promotes specific types of content and agendas, aiming to influence user consciousness and subvert their choices. To target advertisements efficiently, intermediaries collect information about users and utilize tools that allow the intermediary to engage in a new level of refined targeting. For example, Facebook developed the Pixel tool—an interoperable code that

[26] Michal Lavi, *Taking Out of Context*, 31 HARV. J.L. & TECH. 145, 147 (2017).

[27] *See, e.g.*, TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 52–54 (2018) (comparing social networks' community standards regarding sexual content).

[28] *See* ARAL, *supra* note 19, at 203 (expanding on the attention economy).

[29] SIVA VAIDHYANATHAN, ANTISOCIAL MEDIA: HOW FACEBOOK DISCONNECTS US AND UNDERMINES DEMOCRACY 4–7 (2018). *See* Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. 867, 869 (2020); Mark Bergen, *YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant*, BLOOMBERG (April 2, 2019, 11:29 AM), https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant [https://perma.cc/CY83-NRD6].

[30] CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 117 (2017). *See* Michael Wolfowicz, *Examining the Interactive Effects of Personalization Algorithms (The Filter Bubble) on Network Structure (The Echo Chambers) and the Impact on Radical Beliefs*, HEBREW UNIV. OF JERUSALEM, FEDERMANN CYBER CTR. (Oct. 30, 2019), https://csrcl.huji.ac.il/sites/default/files/csrcl/files/cybersadna.pdf [https://perma.cc/HM96-LM6F] (describing an evidence based experiment on a related context, focusing on the function of algorithm in creating a network of connections that form filter bubbles that contribute to radical beliefs**)**.

[31] ARAL, *supra* note 19, at 226 (explaining that algorithms polarize social media users into homogeneous communities and cause automatic herding markets "where people follow the behavior of others" instead of making independent decisions).

collects data to help advertisers track conversions from Facebook ads, optimize those ads, and build a target audience for future ones.[32] Another tool is "Custom Audiences . . . —a matching system pairing one mode of contact with that person's Facebook profile."[33] These "dark ads" are seen only by their narrowly-targeted recipients, unavailable for public scrutiny.[34] Political fake news is seriously impactful and has meaningful social costs when powerful people pay intermediaries to distribute messages among specific target audiences. Fake news used in targeted advertisements leads to false assumptions that can confuse or dissuade the electorate from voting.[35]

Currently, Facebook neither unilaterally bans nor fact-checks political advertisements. Further, the platform unequivocally allows lies in political advertisements.[36] Just before the 2020 election, Facebook "announce[d] a significant set of restrictions designed specifically to protect the integrity of the . . . election cycle, including a flat ban on new political ads in the week before the election."[37] Google outlined restrictions on political ad targeting based on political affiliation.[38] Twitter banned political advertisements with

[32] Christina Newberry, *The Facebook Pixel: What It Is and How to Use It*, HOOTSUITE (Apr. 26, 2021), https://blog.hootsuite.com/facebook-pixel/ [https://perma.cc/K9BR-D6ED].

[33] *What Is a Facebook Custom Audience and How Can They Grow Online Stores?*, BIGCOMMERCE, https://www.bigcommerce.com/ecommerce-answers/what-is-a-facebook-custom-audience/ [https://perma.cc/T538-CGDD]; *see also* FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI 89 (2020).

[34] BENKLER ET AL., *supra* note 22, at 272–75.

[35] Kreiss & Perault, *supra* note 15.

[36] Mike Isaac & Cecilia Kang, *Facebook Says It Won't Back Down from Allowing Lies in Political Ads*, N.Y. TIMES, https://www.nytimes.com/2020/01/09/technology/facebook-political-ads-lies.html [https://perma.cc/MZ2B-D96D] (Sept. 4, 2020). *See also* Omer Kabir, *Facebook Will Not Restrict Political Lies Ahead of Israel's March Election*, CALCALIST (Jan. 27, 2020, 5:18 PM), https://www.calcalistech.com/ctech/articles/0,7340,L-3783290,00.html [https://perma.cc/J8G8-7SKU].

[37] Bhagwat, *supra* note 6, at 2364.

[38] Romm et al., *supra* note 6; Scott Spencer, *An Update on Our Political Ads Policy*, GOOGLE (Nov. 20, 2019), https://blog.google/technology/ads/update-our-political-ads-policy/ [https://perma.cc/RF5M-HYE9]. Google lifted the restriction on political advertisements after the election and enforced ad policies that focus on prohibiting "demonstrably false information that could significantly undermine trust in elections or the democratic process." Megan Graham, *Google Will Lift Its Ban on Political Ads Thursday*,

exceptions mainly applicable to organizations that were not directly speaking about legislative issues.[39]

As technology advances, the dissemination of false information has greater potential to subvert the truth. Deepfake pictures and movies allow even greater manipulation of the truth.[40] Artificial intelligence and machine-learning algorithms combined with facial-mapping software enable the cheap and easy fabrication of content, inserting individual faces into videos without permission.[41] The result is believable videos of people doing and saying things they never did.[42] Liars can easily avoid accountability by claiming that true statements are fake stories. In contrast, truth-tellers can be held as liars.[43] Though intermediaries try to remove deep fakes and other manipulated videos from their platforms, such policies are limited.[44]

CNBC, https://www.cnbc.com/2020/12/09/google-will-lift-its-ban-on-political-ads-thursday.html [https://perma.cc/N9QB-MDWU] (Dec. 9, 2020, 12:52 PM).

[39] Kate Conger, *What Ads Are Political? Twitter Struggles with a Definition*, N.Y. TIMES (Nov. 15, 2019), nyti.ms/2NSLDOh [https://perma.cc/89ET-5WV5] (explaining that what counts as a political advertisement is in the eye of the beholder); Alex Kantrowitz, *Here's the Major Exception to Twitter's Political Ad Ban*, BUZZFEED NEWS (Nov. 11, 2019, 7:02 PM), https://www.buzzfeednews.com/article/alexkantrowitz/heres-the-major-exception-to-twitters-political-ad-ban [https://perma.cc/5D7J-KYJS].

[40] Lili Levi, *Real "Fake News" and Fake "Fake News"*, 16 FIRST AMEND. L. REV. 232, 253 (2017); Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. 1753, 1760 (2019) (explaining that "the emergence of machine learning through neural network methods . . . [increases . . . the] capacity to create false images, videos, and audio."). Generative adversarial networks, known as GANs, can lead to the production of increasingly convincing and nearly impossible to debunk deep fakes. Chesney & Citron, *supra*. Neural networks can also be used for AI creation of news stories that mimic the style and substance of real news stories. *See* Rowan Zellers et al., *Grover: A State-of-the-Art Defense Against Neural Fake News*, GROVER, https://rowanzellers.com/grover/ [https://perma.cc/6HV7-Z2QC].

[41] Chesney & Citron, *Deepfakes: A Looming Crisis for National Security, Democracy and Privacy?*, LAWFARE BLOG (Feb. 21, 2018, 10:00 AM) https://www.lawfareblog.com/deepfakes-looming-crisis-national-security-democracy-and-privacy [https://perma.cc/8RR5-HQAJ].

[42] *Id.*

[43] Chesney & Citron, *supra* note 40, at 1785–86 (describing how the difficulty in separating truth from falsehood provides a "liar's dividend," because anyone can claim that a true story is fake while his lies are the truth).

[44] *Facebook to Remove Deepfake Videos in Run-Up to 2020 U.S. Election*, REUTERS, (Jan. 7, 2020, 1:48 AM), reut.rs/35yOMZa [https://perma.cc/4H23-L9CZ]. However, content would be removed if it "would likely mislead someone into thinking that a subject of the video said words that they did not actually say" and not regarding to a post that tore

In this era, it becomes almost impossible to separate true from false, engage in honest discussions about matters of public importance, and formulate political views without manipulation. Fake news stories can influence voter consciousness, threaten the political security of citizens, erode faith in election results, and even harm the long-term health of democracy and its institutions.[45] Information is disseminated without professional norms, making it difficult for individuals to trust one another.[46] How does the law react to this widespread dissemination of fake news? Should intermediaries operating platforms bear any liability for fake news stories? And if so, when?

Section 230 of the Communications Decency Act ("CDA")[47] reflects the internet exceptionalism approach, as well as U.S. bias toward free speech.[48] It directs that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[49] Courts have interpreted Section 230 to provide broad immunity to internet users and intermediaries that disseminate information created by others.[50] Consequently, lawsuits against intermediaries are usually blocked.[51] Recent debate over intermediaries' obligations regarding organic content and advertisements, in tandem with the attack on intermediaries' moderation practices,[52] calls for reevaluation of the scope of intermediary-based immunity. This

up the original speech and took it out of context. Michael Levenson, *Pelosi Clashes with Facebook and Twitter Over Video Posted by Trump*, N.Y. TIMES (Feb. 8, 2020), https://www.nytimes.com/2020/02/08/us/trump-pelosi-video-state-of-the-union.html [https://perma.cc/WWB3-XU48] (quoting Facebook's policy).

[45] *See* Richard L. Hasen, *Deep Fakes, Bots, and Siloed Justices: American Election Law in a "Post-Truth" World*, 64 ST. LOUIS UNIV. L.J. 535, 539 (2020); Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH. 106, 138 (2019).

[46] ZEYNEP TUFECKI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 40 (2017).

[47] 47 U.S.C. § 230.

[48] JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 78, 146 (2019).

[49] 47 U.S.C. § 230.

[50] *See generally id.*; *see, e.g.*, Barrett v. Rosenthal, 146 P.3d 510, 519, 528–29 (Cal. 2006) (observing that the plain language of Section 230 is evidence that Congress did not intend for an internet user to be treated differently than an internet provider).

[51] *See infra* Part III.A.

[52] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021).

Article focuses on the dissemination of fake news stories as a case study and argues that the immunity regime must be contextualized and nuanced to correspond with intermediaries' actions and resulting effects. Further, this Article seeks to strike a balance between intermediary liability and immunity.[53]

Part I describes the dynamics that lead individuals to disseminate fake news stories. It then outlines a roadmap of the different roles intermediaries play in the dissemination of fake news, including hosting and moderating content, personalizing algorithmic recommendations on organic content, and deploying targeted advertisements for profit. It explains how intermediaries' top-down influence exacerbates dissemination of fake news. Identifying the role that intermediaries play is the first step toward formulating a liability policy.

Part II details the law governing secondary intermediary liability and the concept of internet exceptionalism. Subsequently, it explores the gradual erosion of immunity, Trump's Executive Order on Preventing Online Censorship, and new legislative bills striving to limit intermediary immunity.[54] Finally, it examines the role intermediaries play in fake news dissemination in light of free speech considerations.

Part III argues policymakers should contextualize internet exceptionalism. It targets the exceptions to exceptionalism and outlines a nuanced liability that avoids disproportionate collateral censorship. Additionally, it proposes transparency obligations for intermediaries' moderation activities, requiring algorithmic impact assessments as part of consumer protection regulations.

I. HOW FAKE NEWS STORIES SPREAD

Individuals have diverse motivations for initiating publication of fake news stories. Some are *narrowly self-interested*, aiming to

[53] *See* Balkin, *supra* note 12, at 90.

[54] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021). *See also* S. 4534, 116th Cong. (2019); S. 4066, 116th Cong. (2020).

promote a political candidate by spreading lies about competitors.[55] Others spread fake stories to promote *general interest* and attract more user attention.[56] There are *altruists* who believe conspiracies and publicize them without checking the facts.[57] Finally, there are *malicious propagators* who publish fake stories solely to infringe public interest and inflict harm.[58] While initial publication may have a limited number of recipients, information recipients may then share the false content with others, leading to extensive dissemination and severe harm.

Initiating a fake news story is one thing, but what prompts others to spread it? First, there is a general "truth bias," because people tend to accept what they hear as truthful.[59] However, the nature of the internet's social environment fuels the distribution of fake news stories at minimal cost. Constant internet connection and ongoing communication allows anyone who is connected to share information. "Thus, an idea can spread exponentially and reach a global [audience] at the click of a button."[60] As a fake story circulates, it gains credibility—the more individuals exposed to a particular statement, the more likely they are to perceive and believe it as a known

[55] Michal Lavi, *Publish, Share, Re-Tweet, and Repeat*, 54 U. MICH. J.L. REFORM 441, 450–51 (2021).

[56] *Id.* at 451.

[57] *Id.*

[58] *Id.*; CASS R. SUNSTEIN, ON RUMORS: HOW FALSEHOODS SPREAD, WHY WE BELIEVE THEM, AND WHAT CAN BE DONE 14–15 (2009).

[59] CASS SUNSTEIN, LIARS: FALSEHOODS AND FREE SPEECH IN AN AGE OF DECEPTION 73 (2021).

[60] Lavi, *supra* note 55, at 451. The internet simplifies the dissemination of information and allows sharing to a wide audience at the click of a button. *See* LEE RAINIE & BARRY WELLMAN, NETWORKED: THE NEW SOCIAL OPERATING SYSTEM 67 (2012); DAVID A. POTTS, CYBERLIBEL: INFORMATION WARFARE IN THE 21ST CENTURY? 30 (2011); Jacqueline D. Lipton, *"We, the Paparazzi": Developing a Privacy Paradigm for Digital Video*, 95 IOWA L. REV. 919, 919 (2010); *see generally* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014).

fact.[61] Moreover, lies tend to spread faster than the truth because lies often hold audience attention by inspiring fear and surprise.[62]

Not all fake stories spread as extensively as others; some are only disseminated locally.[63] Why do some fake stories spread widely while others remain limited in reach? In order to provide an answer, sociologists developed models of collective behavior. Mark Granovetter maintains that the key concept of "threshold" explains these processes.[64] His model assumes that information and ideas become more valuable as more individuals accept and adopt them.[65] An individual's threshold for joining an activity is quantified by the proportion of the group the individual would have to see join in on the activity before doing so too.[66] This model assumes that a person's behavior depends on the number of other people already engaging in that particular behavior.[67] "[O]ne's social network has a huge potential to affect one's decisions to adopt and disseminate certain ideas because people respond to the influences and preferences of others."[68]

[61] *See* NICHOLAS DIFONZO & PRASHANT BORDIA, RUMOR PSYCHOLOGY: SOCIAL AND ORGANIZATIONAL APPROACHES 225 (2007); Gordon Pennycook et al., *Prior Exposure Increases Perceived Accuracy of Fake News*, 147 J. EXPERIMENTAL PSYCH. 1865 (2018) (explaining that the more people hear information, the more likely they are to believe it and spread it); Neil Levy, *The Bad News About Fake News*, SOC. EPISTEMOLOGY REV. & REPLY COLLECTIVE, August 2017, at 20 ("[F]ake news is more pernicious than most of us realize, leaving long lasting traces on our beliefs and our behavior even when we consume it know it is fake or when the information it contains is corrected."); Phillips, *supra* note 21 and accompanying text; CASS R. SUNSTEIN, CONSPIRACY THEORIES AND OTHER DANGEROUS IDEAS 25–27 (2014).

[62] Vosoughi et al., *supra* note 19, at 1146; SUNSTEIN, *supra* note 61, at 76.

[63] CHARLES KADUSHIN, UNDERSTANDING SOCIAL NETWORKS–THEORIES, CONCEPTS AND FINDINGS 153 (2011).

[64] *See* Mark Granovetter, *Threshold Models of Collective Behavior*, 83 AM. J. SOCIO. 1420, 1422 (1978) (explaining that "different individuals require different levels of safety" for joining an activity, such as entering a riot, and vary in the benefits they derive from the activity; the crucial concept for describing variation among individuals is that of a "threshold.").

[65] *See id.* at 1424–28.

[66] *Id.* at 1422.

[67] *Id.*

[68] Lavi, *supra* note 55, at 452 (quoting NICHOLAS A. CHRISTAKIS & JAMES H. FOWLER, CONNECTED: THE SURPRISING POWER OF OUR SOCIAL NETWORKS AND HOW THEY SHAPE OUR LIVES 127 (2009)). *See also* Michal Lavi, *Content Providers' Secondary Liability: A*

In addition to a collective threshold, every individual has a personal threshold for adopting and disseminating ideas.[69] Three types of individuals can be abstractly identified. First, individuals who have prior convictions in favor of a new idea or share the same ideology are "*receptives*."[70] Receptives have the lowest threshold and tend to adopt information they receive and subsequently disseminate it.[71] Second is the "*neutrals*," who do not have an inclination in favor of or against an idea.[72] However, if neutrals notice a few people have accepted and disseminated an idea, they may come to accept, join, and disseminate it.[73] Finally, there are the "*skeptics*"—individuals with a prior disposition against certain ideas. Skeptics have a high threshold for accepting and disseminating ideas and need a great deal of information before doing so. "However, once the evidence becomes overwhelming—[and this evidence may include] beliefs . . . shared by many others—[the] skeptics will join others in accepting the idea."[74]

The proliferation of a fake story depends heavily on the type of individual it encounters at the outset.[75] If the story reaches receptives, the neutrals are more likely to reach their threshold, and the skeptics will follow suit and spread the idea further.[76] When an

Social Network Perspective, 26 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 855, 889 (2016).

[69] Granovetter, *supra* note 64 and accompanying text.

[70] *See* Edward Glaeser & Cass R. Sunstein, *Does More Speech Correct Falsehoods*, 43 J. LEGAL STUD. 65, 67 (2014) (explaining that people have "different prior beliefs and hence different degrees of skepticism." Individuals who believe that the messenger is a truth-teller largely have their beliefs buttressed).

[71] *See* SUNSTEIN, *supra* note 58, at 19–20 (explaining that the individual threshold depends on a person's prior disposition regarding the information).

[72] *Id.* at 20.

[73] *See id.*

[74] Michal Lavi, *Evil Nudges*, 21 VAND. J. ENT. & TECH. L. 1, 17 (2018); *see also* Lavi, *supra* note 55, at 453.

[75] *See generally* SUNSTEIN, *supra* note 59; Lavi, *supra* note 74; Lavi, *supra* note 55, at 453; *see, e.g.*, BENKLER ET AL., *supra* note 22 and accompanying text.

[76] *See* SUNSTEIN, *supra* note 59, at 83; Lavi, *supra* note 55, at 453–54. Because individuals influence one another, fake stories can spread through informational cascades. Informational cascades are generated when individuals follow the statements or actions of predecessors and do not express their opposing opinions because they believe their predecessors are right. As a result, the social network fails to obtain important information. *See* Cass R. Sunstein & Reid Hastie, *Four Failures of Deliberating Groups* 2 (Univ. Chi. L. Sch. Pub. L. & Legal Theory Working Paper, Paper No. 215, 2008),

increasing number of people believe a fake news story, it begins to appear credible, influencing others to believe it. Social pressure also pushes people to spread information.[77] Bots are active on many platforms and echo fake stories, exacerbating the likelihood of cascades and fake stories' dissemination.[78] These algorithmic software programs, which run according to programmed instructions, can interact socially with users and enhance trust in online communication.[79] The program creates an impression that many users shared a fake story and triggers human engagement with the content.[80] Consequently, it becomes more likely that individuals will reach their thresholds to believe a fake story and follow the herd.

Many times, fake stories spread to an "influential entity" in a social network, such as a political candidate.[81] If this influential entity accepts and spreads a story, the likelihood increases exponentially that the story will reach a tipping point.[82] This example demonstrates the importance of influential entities, whether an individual or a central website, and the structure of the social network.[83]

https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1125&context=public_law_and_legal_theory [https://perma.cc/4SLD-DXT9]. *See, e.g.*, Matthew J. Salganik et al., *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 SCI. 854, 855 (2006).

[77] In these cases, "people think they know what is right, or what is likely to be right, but they nonetheless go along with the crowd in order to maintain" their status. *See* Sunstein & Hastie, *supra* note 76, at 15. This is the phenomenon of reputation cascades. *Id.*

[78] JARON LANIER, TEN ARGUMENTS FOR DELETING YOUR SOCIAL MEDIA ACCOUNTS RIGHT NOW 55–58 (2018) ("If your extended peer group contains a lot of fake people, calculated to manipulate you, you are likely be influenced without even realizing it.").

[79] Emilio Ferrara et al., *The Rise of Social Bots*, COMMC'NS THE ACM, July 2016, at 99; ARI EZRA WALDMAN, PRIVACY AS TRUST—INFORMATION PRIVACY FOR AN INFORMATION AGE 90, 141 (2018) (expanding on bots that are designed to enable social communication, motivating people to let down their guard against invasions of privacy).

[80] ARAL, *supra* note 19, at 48 ("the early tweeting activity by bots triggers a disproportionate amount of human engagement, creating cascades of fake news, triggered by bots but propagated by humans through the Hype Machine's network.").

[81] *Id.* (explaining that when influential people share content, they can legitimize it and exacerbate its dissemination widely on a social network).

[82] WALDMAN, *supra* note 79, at 146; *see also* MALCOLM GLADWELL, THE TIPPING POINT: HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE 60 (2000) (referring to individuals who possess a great deal of information as "mavens").

[83] *See* BENKLER ET AL., *supra* note 22 and accompanying text.

The proliferation of a story thereby depends heavily on the individuals who encounter its inception.[84] It is difficult, however, to predict these tipping points when ideas are widely spread, as every individual in the network has a different threshold.[85] Changes in a social network's composition, social structures, and the transition path of an idea can significantly alter the likelihood of widespread dissemination.[86]

A. *Roadmap: Intermediaries' Roles in Information Dissemination and the Harm*

Twenty-first century intermediaries are not merely passive conduits; they take on active roles in the dissemination of content and influence the likelihood that individuals will cross their personal thresholds for disseminating fake news stories. This Part maps the roles intermediaries play in withholding or accelerating the dissemination of information: (1) hosting and providing content-neutral tools and interfaces for dissemination; (2) moderating; (3) deploying algorithmically personalized recommendations on organic content; and (4) using targeted advertising to generate profit.

1. Basic Intermediation: Hosting and Providing Content-Neutral Tools and Interfaces for Dissemination

Intermediaries offer platforms for creating content and encourage ongoing engagement with their sites. They utilize technologies and design tools that allow users to sort through vast amounts of information, as well as share various kinds of content. However, users can abuse the platforms to spread lies and fake news.[87] Intermediaries are not passive hosts; they incentivize users to share and disseminate more information because social engagement keeps users

[84] *See id.* at 155, 156 fig.5.6.

[85] *See* Granovetter, *supra* note 64, at 1423 (exemplifying this point through the diffusion of rumors).

[86] CHRISTAKIS & FOWLER, *supra* note 68, at 7 (explaining that social networks and the connections that compose them have dramatic influence over our choices).

[87] *See, e.g.*, Hannah Ritchie, *Read All About It: The Biggest Fake News Stories of 2016*, CNBC, cnbc.com/2016/12/30/read-all-about-it-the-biggest-fake-news-stories-of-2016.html [https://perma.cc/CVJ6-XR9w] (Dec. 30, 2016, 2:04 AM) (discussing a fake news story spread on social media that claimed the pope endorsed Donald Trump).

engaging with the platforms longer.[88] As participation increases, intermediaries earn more revenue from advertisers.[89] Continuing participation allows intermediaries to collect more user information, target personalized advertisements, and maximize profit.[90]

Intermediaries encourage user participation and social sharing by enhancing motivation to spread content, making it easier to share content and triggering users to do so.[91] Although these strategies

[88] ANDREW MARANTZ, ANTI-SOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS AND THE HIJACKING OF THE AMERICAN CONVERSATION 80 (2019) ("Facebook's larger goal, which always went unstated, was not to spread high-quality content; it was to entice more users into spending more time on Facebook."); *see* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 140 (2017).

[89] MARY ANNE FRANKS, THE CULT OF THE CONSTITUTION 171 (2019) ("The more content users voluntarily provide (posts, shares, likes etc.), the more users interact on the platform, and the more companies like Facebook can target users with increasingly personal advertising. If harmful content provided by a user generates a high level of engagement from a large number of users, then the advertising benefit of that post goes up, which means more money in Facebook's pocket.").

[90] This Article will describe this in Part II.I.C. *See generally* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 9 (2019) (coining the term "surveillance capitalism" to describe tracking users' engagement to enhance commercial profits); Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)*, 118 MICH. L. REV. 1073, 1085–86; FRANKS, *supra* note 89; JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 65 (2019) ("Platform-based, massively intermediated environments enable people seeking connection with each other to signal their affinities and inclinations using forms of shorthand—'Like,' 'Follow,' 'Retweet,' and so on—that simultaneously enable data capture and extraction.").

[91] *See* B.J. FOGG, PERSUASIVE TECHNOLOGY: USING COMPUTERS TO CHANGE WHAT WE THINK AND DO 198 (2003) (referring to socio-technical tools for enhancing user motivation and capability to spread content, used by intermediaries to trigger users to spread information); WALDMAN*, supra* note 79, at 90; Cohen, *supra* note 88, at 140; NICHOLAS CARR, THE BIG SWITCH: REWIRING THE WORLD, FROM EDISON TO GOOGLE 154–57 (2009); JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH 72, 102 (2012). This Part describes common intermediary strategies to push users to share more posts, news, and information, regardless of the content shared. Yet, intermediaries can influence the content of information that users share, as the Cambridge Analytica scandal demonstrated. This Part focuses on how intermediaries influence user decisions to share more information, regardless of content. For further discussion of the Cambridge Analytica Scandal, see Sam Meredith, *Here's Everything You Need to Know About the Cambridge Analytica Scandal*, CNBC, https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html [https://perma.cc/TY8R-GE4R], (Mar. 23, 2018, 9:21 AM).

enhance the dissemination of both lies and truths, lies are more likely to spread than truths.[92] The platform's architecture influences decisions to generate and disseminate content.[93] Just a few tweaks in the design of an intermediary's interface can make a huge difference in how it is used and its potential for widespread circulation of ideas. Intermediaries utilize insights gleaned from sociology, psychology, and management that allow them to predict cognitive biases and social dynamics, deploy new socio-technical systems, and influence the flow of information.[94] Much like the gaming industry, design and technology turn the use of social media into an inherent need.[95] This causes users to become addicted to the platform, keeping them on the website.[96] To make their platforms "sticky" and enhance dissemination, intermediaries "[organize] everything around friending, clicking, retweeting, . . . responding,"[97] and exhibiting to others.[98] Pictures, names, and other informal touches give the impression that online contacts are well-known friends. This choice architecture and framing not only increases users' addiction to a

[92] Vosoughi et al., *supra* note 19 (explaining how researchers revealed that fake stories are disseminated significantly "farther, faster, deeper, and more broadly" than true ones).

[93] *See* FOGG, *supra* note 91, at 5.

[94] TUROW, *supra* note 91, at 74; Lavi, *supra* note 55, at 461.

[95] ZUBOFF, *supra* note 90, at 466 (explaining that "[j]ust as ordinary consumers can become compulsive gamblers at the hands of the gaming [industry]," behavioral technology can draw ordinary young people into an unprecedented vortex of social information); Lavi, *supra* note 55, at 451.

[96] ZUBOFF, *supra* note 90, at 466.

[97] BERNARD E. HARCOURT, EXPOSED: DESIRE AND DISOBEDIENCE IN THE DIGITAL AGE 41 (2015).

[98] *Id.* at 41, 90; Daniel Susser et al., *Online Manipulation: Hidden Influences in a Digital World,* 4 GEO. L. TECH. REV. 1, 29–30 (2019) ("Both the information [we] knowingly disseminate about [ourselves . . . when we] visit websites, make online purchases, and post photographs and videos on social media[,] and the information [we] unwittingly provide . . . as] those websites record data about how long [we] spend browsing them, where [we] are when [we] access them, and which advertisements [we] click on[,] reveals a great deal about who [we . . . are], what interests [us], and what [we] find amusing, tempting, and off-putting.").

platform,[99] but also the likelihood that an individual will reach his threshold and share information he would not otherwise share.[100]

Moreover, intermediaries make it easier than ever to disseminate any kind of content. For example, "share" and "re-tweet" buttons make re-posting content incredibly easy by enabling users to share content at the click of a button.[101] Thus, users need not go through the more cumbersome copy-and-paste process to spread content. Due to the low cost of sharing information, it is more likely that individuals will cross their thresholds and join others already engaged in information dissemination.[102] Simplifying "the re-posting process" encourages users to share information intuitively and instinctively, bypassing reflective thinking about the consequences of dissemination.[103] This choice architecture engineers social behavior and influences decisions to share information.[104]

[99] HARCOURT, *supra* note 97, at 122 (referring to the collection of information using the metaphor, "the glass mirror"); LANIER, *supra* note 78, at 21–23, 29 ("[A]ddiction is a big part of the reason why so many of us accept being spied on and manipulated by our information technology."); *see also* Katie Mettler, *A Lawmaker Wants to End 'Social Media Addiction' by Killing Features That Enable Mindless Scrolling*, WASH. POST (July 30, 2019), wapo.st/2KBQ3X5 [http://perma.cc/2WF6-WXU2]; WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 198 (2018) (expanding on architecture that causes users to become addicted to engagement); ZUBOFF, *supra* note 90, at 466 and accompanying text.

[100] *See* Ari Ezra Waldman, *Privacy, Sharing, and Trust: The Facebook Study*, 67 CASE W. RSRV. L. REV. 193, 203 (2016); Ari Ezra Waldman, *Safe Social Spaces*, 96 WASH. U. L. REV. 1537, 1565 (2019); Ari Ezra Waldman, *Cognitive Biases, Dark Patterns, and the "Privacy Paradox"*, 31 CURRENT OPINION IN PSYCH. 105, 108–09 (2020) (explaining that platforms deliberately create social cues to encourage sharing); HARCOURT, *supra* note 97, at 86, 99; NICHOLAS CARR, THE GLASS CAGE: AUTOMATION AND US 177–82 (2014); James Grimmelmann, *Accidental Privacy Spills*, 12 J. INTERNET L. 3, 6 (2008); *see generally* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013).

[101] Lavi, *supra* note 55, at 464.

[102] SUNSTEIN, *supra* note 30, at 108.

[103] DANIEL KAHNEMAN, THINKING FAST AND SLOW 237 (2011) (explaining the two systems of thinking, or modes of thought: intuitive thinking ("System 1") and deliberative analytical thinking ("System 2")).

[104] BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 70, 235 (2018) ("The smart social media environment that has emerged in the past decade of which Facebook is an important part—encourages people to accept what [is] presented to them without pushing for reflection or deliberation.").

2. Moderation

Intermediaries shape the flow of information, influencing what is viewed, valued, and disseminated. Moderation of users' content is one way platforms shape public discourse. It "promotes adherence to the platforms' terms of use statements, site guidelines, and legal regimes. It is a key part of the production chain of commercial sites and social media platforms."[105] Professor Tarleton Gillespie posits that intermediaries must moderate content; in fact, he demonstrates that moderation is a fundamental aspect of any platform.[106] Many interviews with moderators show that moderation is necessary for proper operation of online platforms, and intermediaries recognize that moderation is a critical part of their production chain.[107] "Social media companies often regulate speech in many different ways, using different tools."[108] Companies govern speech, enforce policies and terms of services, and moderate harmful content,[109] such as fake news and incitement, even though they are not obligated to do so. They can moderate content before it is published on their sites (ex-ante moderation) or after (ex-post moderation).[110] Moderation may be reactive, such as when users send notice to moderators of inappropriate content, or proactive, such as when moderators seek out published content for removal.[111] It can be done automatically by software or manually by humans.[112] Moderators that operate without sufficient transparency can remove or obscure content, making it

[105] Michal Lavi, *Do Platforms Kill?*, 43 HARV. J.L. & PUB. POL'Y 477, 496 (2020) (citing SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOW OF SOCIAL MEDIA 71 (2019)).

[106] GILLESPIE, *supra* note 27, at 5–6.

[107] ROBERTS, *supra* note 105, at 203.

[108] Lavi, *supra* note 105, at 497; *see also* Eric Goldman, *Content Moderation Remedies*, MICH. TECH. L. REV. (forthcoming 2021) (manuscript at 1) (on file with author) (reviewing the range of options to "redress content or accounts that violate the applicable rules").

[109] *See* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1625–30 (2018) (explaining that the reasons for moderation are corporate responsibility and economics).

[110] *Id.* at 1635.

[111] *Id.*

[112] *Id.*

less visible.[113] They can also suspend accounts.[114] "Intermediaries can and do moderate content" to enforce violations of the platforms' terms of services and community guidelines and to mitigate the effects of harmful content in various contexts, such as incitement, defamation, and fake stories.[115] However, intermediaries' approaches toward moderation are inconsistent within the given platform[116] and differ among social media sites.[117] The diverging attitudes of Twitter and Facebook toward Trump's tweets[118] exemplifies such differences. Moderation can influence the visibility of fake news stories and other related harmful content, withhold dissemination, accelerate dissemination, and influence the likelihood of users noticing and believing it.

B. Algorithmically Personalized Recommendations on Organic Content

Intermediaries can promote specific content in users' newsfeeds via algorithmic recommendations on organic user content.[119] They generally optimize relevant content to deliver to users, improving the experience and enhancing engagement.[120] This practice is often

[113] *See, e.g.*, Hern, *supra* note 25; Niva Elkin-Koren & Maayan Perel, *Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law*, *in* OXFORD HANDBOOK ONLINE INTERMEDIARY LIABILITY 669, 674 (Giancarlo Frosio ed., 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3542992 [https://perma.cc/P7E2-WL98].

[114] *See* Goldman, *supra* note 108, at 37.

[115] Lavi, *supra* note 105, at 497 (describing moderation practices regarding incitement of terror).

[116] GILLESPIE, *supra* note 27, at 117 ("Because this work is distributed among different labor forces, because it is unavailable to public or regulatory scrutiny, and because it is performed under high-pressure conditions, there is a great deal of room for slippage, distortion, and failure.").

[117] *Id.* at 20 ("Platforms vary, in ways that matter both for the influence they can assert over users and for how they should be governed."); *see also* Lavi, *supra* note 105, at 498 (giving a contextually related example: Twitter and Facebook have different attitudes toward moderation of incitement of terror).

[118] *See, e.g.*, Spangler, *supra* note 1.

[119] *See, e.g.*, FRISCHMANN & SELINGER, *supra* note 104, at 117–18 (describing Facebook's experiment in which the company displayed negative posts by "friends," and omitted positive posts).

[120] COHEN, *supra* note 90, at 85 ("Platform-based providers of search, and content aggregation, and social networking services operate at the intersection of behavioral microtargeting and content optimization for engagement.").

problematic,[121] and can result in recommendations for fake news stories and conspiracy theories presented directly to susceptible users.[122]

Although it may appear as if the system operates without human intervention, the algorithm's operation depends on the programmer's discretion.[123] The system's designers can limit algorithmic functions.[124] Intermediaries can prioritize content according to users' characteristics and activities. However, the intermediary can also preference an algorithm programmed without content neutrality and promote specific types of content or agendas according to its own strategic preferences. For example, according to testimony from former Facebook product manager Frances Haugen in front of the Senate committee, Facebook is aware that its algorithm promotes harmful content, and the company still avoids deploying counter-measures.[125] Moreover, the intermediary can tinker with its algorithm to decrease the visibility of specific content or increase the exposure of other items, thereby influencing the likelihood that users will further disseminate the information.[126] For example, Google tinkers with search results to favor specific businesses.[127]

[121] VAIDHYANATHAN, *supra* note 29, at 54–55 (explaining that what makes Facebook good also makes it bad, and demonstrating that Facebook has grown into the most reckless and irresponsible system in the commercial world).

[122] Manheim & Kaplan, *supra* note 45, at 147.

[123] Ari Ezra Waldman*, Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 615 (2019) (explaining that the operation of the algorithm is part of the neoliberal managerial project).

[124] Philip S. Thomas et al., *Preventing Undesirable Behavior of Intelligent Machines*, 366 SCI. 999, 1003 (2019); Lauren E. Willis, *Deception by Design*, 34 HARV. J.L. & TECH. 115, 181 (2020).

[125] Ryan Mac & Cecilia Kang, *Whistle-Blower Says Facebook 'Chooses Profits Over Safety'*, N.Y. TIMES, https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html [https://perma.cc/NSK5-GJT4] (Oct. 27, 2021, 5:51 PM).

[126] *See generally* Omer Tene & Jules Polonetsky, *Taming the Golem: Challenges of Ethical Algorithmic Decision-Making*, 19 N.C. J.L. & TECH. 125, 137–38 (2017) (differentiating between "policy-neutral algorithms" that can, in some cases, reflect existing entrenched societal biases and historic inequalities, and in contrast, "policy-directed algorithms" that are purposefully designed to advance a predefined policy agenda).

[127] Google's algorithms are subject to regular "tinkering" by executives and engineers to generate specific search results, including algorithms affecting topics such as vaccinations and autism. *See* Kirsten Grind et al., *How Google Interferes with Its Search Algorithms and Changes Your Results*, WALL ST. J. (Nov. 15, 2019, 8:15 AM), https://www.wsj.com/

Similarly, intermediaries can use algorithms to favor specific candidates in elections, systematically prefer specific businesses, and spread algorithmic propaganda to influence users in favor of specific viewpoints.[128]

Intermediaries are engineered to promote items that generate reactions and elicit strong emotions through algorithmic recommendations. This includes fake news stories and extremist content.[129] They influence the type of content users see and increase the likelihood of reaching users' thresholds to pass along content. The Facebook cognition experiment demonstrates this: Facebook displayed only negative posts with negative words to some users, while displaying positive posts to others. Users exposed only to negative posts created similar posts and shared them at higher rates than other types of content.[130] Users exposed only to positive posts disseminated more positive posts.

Even when intermediaries use algorithms not aimed at promoting specific types of content, the algorithm is never absolutely neutral.[131] It personalizes recommendations users see on their news

articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753 [https://perma.cc/YAW5-GJ3S].

[128] *See generally* LANIER, *supra* note 78, at 81–92.

[129] Jonah Berger & Katherine L. Milkman, *What Makes Online Content Viral?*, 49 J. MKTG. RES. 192, 193–205 (2012); MARANTZ, supra note 88, at 79 ("Content that evokes high-arousal emotion is more likely to be shared . . . ."); *see id.* at 118 ("The algorithms were not designed to gauge whether an idea was true or false, prosocial or antisocial; they were designed to measure whether a meme was causing a spike of activating emotions in a large number of people."); VAIDHYANATHAN, *supra* note 29, at 5–9 (describing how Facebook develops algorithms that favor highly charged content and depends on self-serving advertising systems that precisely target ads using massive surveillance and personal dossiers); LANIER, *supra* note 78, at 23, 120–21; Citron, *supra* note 90; ZUBOFF, *supra* note 90, at 301; Mark Bergen, *YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant*, BLOOMBERG (Apr. 2, 2019, 11:29 AM), bloom.bg/36UCOLi [https://perma.cc/AFX6-QRXK].

[130] Adam D. I. Kramer et al., *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PNAS 8788, 8788 (2014); FRISCHMANN & SELINGER, *supra* note 104, at 117–18 (describing the impact of Facebook's cognition experiment on user emotions); *see also* Manheim & Kaplan, *supra* note 45, at 147 ("Google's YouTube also profits nicely from fake news. Its 'recommendation algorithm' serves 'up next' video thumbnails that its AI program determines will be of interest to each of its 1.5 billion users.").

[131] FRANKS, *supra* note 89, at 186 ("[W]hile algorithms are built on data, they also 'optimize output to parameters the company chooses, crucially, under conditions also

feeds based on collected information, social network engagement, social cues—such as clicks on content, "likes," and "shares"—and past activity on the platform.[132] Intermediaries can also characterize users by their social relations and friends within a social network. Such information allows artificial intelligence algorithms to show users personalized recommendations for relevant content. By "systemization of the personal,"[133] intermediaries influence and even control with whom users connect, what they see online, and the visibility of specific content.[134] Thus, intermediaries selectively influence the content users see on their newsfeeds and do not present content chronologically.[135]

Personalizing content does not offer equal choice to all users. The intermediary's algorithm determines what recommendations and content will be available to whom.[136] Thus, different people see different content and have varied online experiences.[137] Personalized prioritization of content can result in socio-technological engineering and have self-reinforcing power.[138] Algorithmically

shaped by the company' . . . .Algorithms, in other words, are human choices all the way down.").

[132] Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2027 (2018) ("The creation of personalized feeds is inevitably content based—social media sites have to decide what content is likely to be most interesting to their end users.").

[133] Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1021 (2014).

[134] For a related context of discrimination, see Karen Levy & Solon Barocas, *Designing Against Discrimination in Online Markets*, 32 BERKELEY TECH. L.J. 1183, 1183 (2017) ("[P]latforms necessarily exercise a great deal of control over how users' encounters are structured—including who is matched with whom for various forms of exchange.").

[135] Balkin, *supra* note 132 and accompanying text.

[136] ARAL, *supra* note 19, at 211 (expanding on optimization algorithms); *id.* at 220 ("algorithmic emphasis on trending make the New Social Age rife with inequality").

[137] Olivier Sylvain, *Discriminatory Designs on User Data*, COLUMBIA U. KNIGHT FIRST AMEND. INST. (Apr. 1, 2018), https://knightcolumbia.org/content/discriminatory-designs-user-data [https://perma.cc/PQ85-RRG2] ("[M]any intermediaries analyze, sort, and repurpose the user content they elicit. Facebook and Twitter, for example, employ software to make meaning out of their users' 'reactions,' search terms, and browsing activity in order to curate the content of each user's individual feed, personalized advertisements, and recommendations about 'who to follow.'").

[138] *Id.*; *see also* ZUBOFF, *supra* note 90, at 19 ("New automated protocols are designed to influence and modify human behavior at scale as the means of production is subordinated to a new and more complex means of behavior modification."); LANIER, *supra* note 78, at 28–29 (referring to this algorithmic propaganda as "BUMMER"—Behavior of Users Modified and Made into Empire for Rent).

personalized recommendations can influence future users' choices and the likelihood users will change their minds.[139] Because algorithmic recommendations are tailored to the user's characteristics, they usually reinforce pre-existing beliefs. In other words, the user is "receptive" to information that confirms prior dispositions and more likely to accept such information as true.[140] Algorithmic recommendations do not end with influencing individual beliefs. Rather, they affect dissemination throughout the network, resulting in a feedback loop that reinforces the individual's cluster of connections who share similar characteristics.[141] As the algorithm narrows information available to a user and his social connections, it helps create filters applied to individuals echoing similar opinions.[142] As a result, the marketplace of ideas is hampered.

[139] Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, 23 MICH. TECH. L. REV. 59, 60–61 (2018); *see also* ZUBOFF, *supra* note 90, at 20; Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. U. L. REV. 1353, 1360 (2018) ("[D]igital expressive opportunities are neither limitless nor uniform.").

[140] *See* COHEN, *supra* note 90, at 85 ("[S]ocial networking providers like Facebook and microblogging platforms like Twitter function as de facto aggregators for a wide range of content and deliver feeds optimized to everything that is known or inferred about particular users' opinions and beliefs.") For more information on the confirmation bias, see SUNSTEIN, *supra* note 30, at 122–24; BENKLER ET AL., *supra* note 22, at 76 (explaining that individuals "look for media outlets and politicians that will inform them as best as possible without suffering too much cognitive discomfort.").

[141] *See* Julie E. Cohen, *The Emergent Limbic Media System*, *in* LIFE AND THE LAW IN THE ERA OF DATA-DRIVEN AGENCY 61, 72 (Mireille Hildebrandt & Kieron O'Hara eds., 2020); Julie E. Cohen, *Internet Utopianism and the Practical Inevitability of Law*, 18 DUKE L. & TECH. REV. 85, 88 (2019) (referring to the feedback loop caused by algorithmic recommendations and targeting, noting "[a]lgorithmic processes optimized to boost click-through rates and prompt social sharing heighten the volatility of online interactions, and surveillant assemblages designed to enhance capabilities for content targeting and behavioral marketing create powerful—and easily weaponized—stimulus-response feedback loops.").

[142] *See* SUNSTEIN, *supra* note 30, at 98–136. It should be noted that Eli Pariser was one of the first to warn that algorithms show links that users are more likely to click. ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 35–48 (2011); *see also* MARANTZ, *supra* note 88, at 76 (explaining that filter bubbles are not a bug, "but a central feature of social media. It is hard to see how [. . . it] could flourish without [them]."); PARISER, *supra*, at 157–58 ("[E]ach social network developed its own set of content-sorting algorithms, many of which, despite the good intentions of the engineers who built them, would start to function as filter bubbles or radicalization engines.").

C. Targeting Advertisements for Profit

Facebook will run any "political" ad you want, even if it's a lie. And they'll even help you micro-target those lies to their users for maximum effect. Under this twisted logic, if Facebook were around in the 1930s, it would have allowed Hitler to post [thirty]-second ads on his "solution" to the "Jewish problem."[143]

Intermediaries directly profit from targeting advertisements.[144] This type of influence on the information flow is different from algorithmically personalized recommendations on organic content. Whereas recommendations regarding organic content provide users with relevant content and enhance engagement by using policy-neutral algorithms, targeted advertisements aim to promote the advertisers' agendas. They use policy-directed algorithms and tools biased in favor of the advertisers' agenda and are anything but neutral to content.[145] The social media advertising ecosystem is a persuasion market.[146] In this capacity, intermediaries develop special strategies for refined targeting and create a different context for the information. By leveraging the enormous amount of user data they collect, analyzing it, and developing cutting edge micro-targeting tools, intermediaries offer advertisers the opportunity to display "the right ad, to the right person, at the right time,"[147] thereby influencing user consciousness and subverting user decision-making. The following Part describes these stages.

[143] Cohen, *supra* note 24.

[144] ARAL, *supra* note 19, at 203 ("Platforms like Facebook, Twitter and YouTube provide connections, communication and content to get consumers' attention. They then sell that attention to brands, governments, and politicians who want to change people's perceptions, opinion and behaviors with ads.").

[145] Tene & Polonetsky, *supra* note 126, at 137–39 (differentiating between "policy-neutral algorithms" and "policy directed algorithms").

[146] ARAL, *supra* note 19, at 133 ("[S]ocial media advertising ecosystem is a persuasion market. Brands, governments and political campaigns invest in it to persuade us to change our behavior, from how we vote to what products we buy.").

[147] *People-Based Marketing: Thinking People-First Planning and Measurement*, FACEBOOK IQ (Sept. 13, 2017), https://www.facebook.com/business/news/insights/the-future-of-marketing-people-based-planning-and-measurement [https://perma.cc/8UV8-LLJV]; *see also* COHEN, *supra* note 90, at 180 ("Targeted advertising can ensure that consumers see only certain options, and cutting-edge behavioral microtargeting techniques that identify points of vulnerability can be used to shape and refine targeting strategies.").

1. Data Collection

Intermediaries seduce users into sharing more information by using interfaces designed to encourage sharing, subsequently exposing users to robust collection of personal data.[148] For example, platforms can offer personality questionnaires and draw information on users' traits for profiling.[149] In addition to information users willfully disseminate while engaging with others, intermediaries are constantly collecting user data incidental to everyday user activity without the individual's awareness.[150] The rapid move into a world dominated by the Internet of Things ("IoT") merges individuals' online activities with their offline activities and enables companies to collect data in domains traditionally perceived as offline realms.[151] "Every minute of every day, everywhere on the planet, dozens of companies . . . are logging the movements of millions of people with mobile phones and storing the information in gigantic data files,"[152] providing intermediaries with troves of user data. [153]

As the recent Facebook leak demonstrated, companies can collect data for one purpose and share it with third parties for other

[148] *See infra* Part II.A.

[149] *See* Meredith*, supra* note 91.

[150] *See* ZUBOFF, *supra* note 90, at 80–81; COHEN, *supra* note 90, at 42. Ninety-two percent of websites have embedded Google trackers, so that the company knows about every place a person visits on the internet—whether or not he has a Google account or uses any Google services. *See* Ibrahim Altaweel et al., *Web Privacy Census*, TECH. SCI. (Dec. 14, 2015), https://techscience.org/a/2015121502/ [https://perma.cc/UB69-6K8C]; ARAL, *supra* note 19, at 206 (explaining that microtargeting models are powered by reams of personal data about consumers' demographics, behaviors, preferences, and psychological profiles).

[151] GILAD ROSNER & ERIN KENNEALLY, UNIV. CAL. BERKELEY CTR. FOR LONG-TERM CYBERSECURITY, PRIVACY AND THE INTERNET OF THINGS 7 (2018) ("As the Internet of Things expands, this type of granular data collection is moving into domains that have traditionally been considered 'offline.'").

[152] Stuart A. Thompson & Charlie Warzel, *Twelve Million Phones, One Dataset, Zero Privacy*, N.Y. TIMES (Dec. 19, 2019), nyti.ms/2Zfby6E [https://perma.cc/K3D7-TBKV]; *see* COHEN, *supra* note 90, at 57 ("[S]ubsequent continuing extensions of surveillance capability have been more deliberate. The primary vehicles for those extensions have been the marketplace shifts toward smart mobile devices, wearable computing, and the internet of things.").

[153] ROSNER & KENNEALLY, *supra* note 151, at 5. This scale of collection is made possible through smart connected devices, such as wearables, digital assistants, smart speakers, fitness trackers, and other gadgets that include sensors that sense and monitor our every utterance.

purposes.[154] A person can share information with an application that is later transferred to Facebook.[155] In many cases, users consent to data collection without understanding the implications. This can be attributed to dark patterns in the platform's architecture that intentionally confuse users into clicking "I agree."[156] In other cases, individuals have no choice but to consent, because there is no equivalent alternative to the service.[157]

2. Analyzing and Profiling

Intermediaries translate raw data they collect into behavioral insights about users and third parties.[158] They collect data from a variety of sources, and users allow them to identify and extract unpredictable value from such data by exploiting new capabilities in data

[154] Sebastian Klovig Skelton & Bill Goodwin, *Lawmakers Study Leaked Facebook Documents Made Public Today*, COMPUT. WKLY. (Nov. 6, 2019, 1:21 PM), https://www.computerweekly.com/news/252473540/Lawmakers-study-leaked-Facebook-documents-made-public-today [https://perma.cc/E956-8ATG] (revealing the document leak and explaining that "Facebook planned to use its Android app to match users' location data with mobile-phone base station IDs to deliver 'location-aware' products without users' consent." Facebook also gave preference to certain deals to partners who shared their user data with Facebook.); *see also Facebook Sold a Rival-Squashing Move as Privacy Policy, Documents Reveal*, GUARDIAN (Nov. 6, 2019, 01:48 PM), https://www.theguardian.com/us-news/2019/nov/06/facebook-privacy-switcharoo-plan-emails [https://perma.cc/LS54-9DPV].

[155] Calo, *supra* note 133, at 1004; *see, e.g.*, Sam Schechner & Mark Secada, *You Give Apps Sensitive Personal Information. Then They Tell Facebook.*, WALL ST. J. (Feb. 22, 2019, 11:07 AM), on.wsj.com/2HsnY40 [https://perma.cc/GL83-YW8H].

[156] Waldman, *Cognitive Biases, Dark Patterns, and the "Privacy Paradox"*, *supra* note 100, at 107 ("designers intentionally make it difficult for users to effectuate their privacy preferences."); *id.* at 108 ("Dark patterns can hide disclosure dangers while simultaneously highlighting the powerful social cues to share."); Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43, 43 (2021) ("Dark patterns are user interfaces whose designers knowingly confuse users, make it difficult for users to express their actual preferences, or manipulate users into taking certain actions. They typically prompt users to rely on System 1 decision-making rather than more deliberate System 2 processes.").

[157] CARISSA VELIZ: PRIVACY IS POWER: WHY AND HOW YOU SHOULD TAKE BACK CONTROL OF YOUR DATA 39 (2020) (explaining that during COVID-19 lockdowns individuals were in fact forced to agree to Zoom's terms of service in order to work and to allow their children to attend distance learning; in fact, the service became indispensable in order to be full participants in society).

[158] Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11, 17 (2020) ("As digital companies know more about a given person, they can also know more about other people who are similar to that person or are connected to that person.").

analysis.[159] Complex algorithms mine the information, integrate it, find connections and correlations between data items, identify patterns, and draw conclusions about individuals.[160] Analyzing "likes" on Facebook allows intermediaries to evaluate a wide range of personality traits, emotional states,[161] and psychographic traits[162] and discover facts about users—even facts users never meant to share with anyone.[163]

Beyond obtaining knowledge about a user's present emotional state and reactions, processing data on user behavior can forecast future feelings and thoughts.[164] The result is not just a feedback but also a feed-forward by looking backwards.[165] For example, Cambridge Analytica used information about users' personality traits drawn from personality questionnaires to develop a model for

[159] COHEN, *supra* note 90, at 56 ("'Big Data,' was fast-evolving group of techniques for converting voluminous, heterogeneous flows of physical, transactional, and behavioral information about people."); *see also id.* at 66 ("After personal data have been cultivated and harvested, they are processed to generate patterns and predictions about data subjects' preferences and behavior."); Max N. Helveston, *Consumer Protection in the Age of Big Data*, 93 WASH. U. L. REV. 859, 867 (2016) (articulating the distinguishing characteristics of Big Data analytics: volume, velocity, and variety); Fred H. Cate & Viktor Mayer-Schönberger, *Notice and Consent in a World of Big Data*, 3 INT'L DATA PRIV. L. 67, 69 (2013) (discussing the ubiquity of data collection and technological developments that expand the ability to analyze, identify, and extract new value from seemingly worthless data).

[160] VIKTOR MAYER SCHÖNBERGER & THOMAS RAMGE, REINVENTING CAPITALISM IN THE AGE OF BIG DATA 77–78 (2018).

[161] *See* Michal Kosinski et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 PNAS 5802, 5802 (2013); Youyou Wu et al., *Computer-Based Personality Judgments Are More Accurate Than Those Made by Humans*, 112 PNAS 1036, 1037–38 (2015).

[162] Psychographic profiles were at the core of the Cambridge Analytica Scandal. *See* Hannes Grassegger & Mikael Krogerus, *The Data That Turned the World Upside Down*, VICE (Jan. 28, 2017, 9:15 AM), https://www.vice.com/en/article/mg9vvn/how-our-likes-helped-trump-win [https://perma.cc/HZC2-MX29]; VAIDHYANATHAN, *supra* note 29, at 150–54; *see generally* Terrell McSweeny, *Psychographics, Predictive Analytics, Artificial Intelligence & Bots: Is the FTC Keeping Pace?*, 2 GEO. L. TECH. REV. 514 (2018).

[163] ZUBOFF, *supra* note 90, at 274–77; Gregory Park et al., *Automatic Personality Assessment Through Social Media Language*, 108 J. PERSONALITY & SOC. PSYCH., 934, 943–44 (2015).

[164] ZUBOFF, *supra* note 90, at 95 (referring to data on the behavior of technology users as "behavioral surplus").

[165] HARCOURT, *supra* note 97, at 145–46.

predicting voter behavior and used it to target political messages.[166] The more data intermediaries collect, the more accurate their predictive algorithms are. More predictive algorithms result in intermediaries' powerful ability to influence users through digital advertising.[167]

3. Developing Targeting Tools

Intermediaries build powerful tools for political and commercial campaigns.[168] They can target advertisements to voters based on data from multiple sources. *First,* intermediaries can use data they collect as described. *Second*, they can use data from third-party companies.[169] *Third,* they can use data shared by advertisers. Analyzing such data allows intermediaries to assign attributes to individual users, define specific target audiences for advertisements, and narrow distribution to the audience most likely to respond.[170]

Intermediaries also develop interfaces that make it easier for advertisers to collect user data and refine their advertisements and potential target audiences.[171] Facebook offers advertising tools for collecting information and provides a vast array of targeting options.[172] The Pixel tool serves as a good example. This code can be operationalized in every website, collecting data to help advertisers track

[166] VAIDHYANATHAN, *supra* note 29, at 155.

[167] Balkin, *supra* note 12, at 84; Kim, *supra* note 29, at 878 (explaining that mass data collection and analysis makes thousands of user attributes available for advertisers to refine their target audiences).

[168] BENKLER et. al., *supra* note 22, at 271–75.

[169] For example, until the Cambridge Analytica scandal, Facebook cooperated with data broker companies like Experian and Acxiom to use their data for more accurate ad targeting. Kurt Wagner*, Facebook Is Cutting Third-Party Data Providers Out of Ad Targeting to Clean Up Its Act*, VOX (Mar. 28, 2018, 6:11 PM), https://www.vox.com/2018/3/28/17174098/facebook-data-advertising-targeting-change-experian-acxiom [https://perma.cc/6QAY-KSP3].

[170] Kim, *supra* note 29, at 878 (explaining that the collection of data allows intermediaries to target advertisements based on geographic location, interest, affiliations, or behaviors).

[171] Kreiss & Perault, *supra* note 15 ("Facebook allows advertisers to bring their own data to their platforms for targeting purposes, and Twitter has similar tools for commercial ads."). *See* MARANTZ, *supra* note 88, at 78 (discussing companies such as Upworthy that specialize in creating clickable and sharable headlines and test them against one another algorithmically to determine which is most popular).

[172] ARAL, *supra* note 19, at 207 (explaining that microtargeting can depend on demographics, behavior, preferences, and psychological profiles).

conversions from Facebook advertisements, optimize those ads, and build target audiences for future ads.[173] It works by placing and triggering cookies that track users as they interact with the advertisers' websites and Facebook ads.[174] This tool allows for data collection and audience targeting through different parameters.[175] For example, the "Lookalike Audience" tool[176] allows advertisers to provide Facebook with information about an existing group—the source audience—which represents its target audience and serves as the basis for targeting. Facebook also provides "Custom Audiences . . . —a matching system pairing one mode of contact with that person's Facebook profile," that allows businesses to interact with relevant users across multiple channels.[177] In most cases, businesses can expect thirty to seventy percent of their contacts to have matching profiles on the platform; Custom Audiences can thereby reach highly targeted individuals.[178]

Likewise, Twitter developed targeting tools based on users' spoken languages, genders, interests, followers of relevant accounts, and devices used.[179] Behavioral targeting is then based on users' activity patterns.[180] Moreover, intermediaries provide interfaces that allow ad campaigns to measure responses to advertisements[181] and

[173] Newberry, supra note 32.

[174] *Id.*

[175] ARAL, *supra* note 19, at 207; Rebecca Uliasz, *"Optimize User Experience": Optimization Techniques and the Simulation of Life, from the Model to the Algorithm*, 21 REV. COMMC'N 129, 137 (2021) ("The pixel permits an advertiser to track, organize, and interpret information about user behaviors on a webpage to target potential customers. Audience data are algorithmically processed by Facebook internally to achieve different optimization aims, such as increasing conversions on a specific ad or maximizing high-value purchases.").

[176] *About Lookalike Audiences*, FACEBOOK BUS., https://www.facebook.com/business/help/164749007013531?id=401668390442328 [https://perma.cc/WS6S-TVNE]; Kim, *supra* note 29, at 879.

[177] *See What Is a Facebook Custom Audience and How Can They Grow Online Stores?*, *supra* note 33.

[178] *Id.*

[179] *See Twitter Ads Targeting*, TWITTER BUS., https://business.twitter.com/en/advertising/targeting.html [https://perma.cc/B3U2-UPC9].

[180] *See id.*

[181] Micah L. Berman, *Manipulative Marketing and the First Amendment*, 103 GEO. L.J. 497, 518 (2015) (explaining that neuro-marketing specialists measure the brain's response to marketing stimuli in real time, allowing companies to determine "individuals' emotional

refine their targeting. These interfaces make it possible to quickly evaluate how well different versions of the same message elicit engagement in the target audience and increase the advertisements' relevance.[182] Intermediaries can experiment with levels of influence, assess feedback, and select the most effective tool.[183]

Targeting tools enable intermediaries to identify specific voters, geographic regions, and demographic segments[184] based on users' personal data.[185] Due to accurate targeting, only the narrowly tailored, intended recipients see "dark ads," making these ads unavailable for public scrutiny.[186] Therefore, it becomes more difficult for watchdogs such as journalists and civil society organizations to detect these advertisements and alert the public to fake news, including politicians' lies.

4. Strategies of Targeting

Targeting susceptible users is only part of the story. The influence strategies that intermediaries utilize are beyond persuasion,[187]

responses to brands and brand preferences, even when the individual may be unaware of the brand's effect on his subconscious decision making.").

[182] *See, e.g.*, *Facebook for Business: Make Smarter Business Decisions with Actionable Insights.*, FACEBOOK, https://www.facebook.com/business/measurement [https://perma.cc/DYL5-HC87] (providing a service that includes A/B testing to compare versions of a single variable in advertisements).

[183] Tal Z. Zarsky, *Privacy and Manipulation in the Digital Age*, 20 THEORETICAL INQUIRIES L. 157, 170 (2019).

[184] Platforms can change ad targeting to avoid some legal violations related to targeting, such as discrimination. However, the change is aimed at the segments of targeting and not the content of the advertisements and strategies of targeting, and therefore the problem of accurate targeting of fake news stories remains. *See* Kim, *supra* note 29, at 878 ("After several lawsuits alleged these tools could be used to discriminate, Facebook agreed in March 2019 to a settlement restricting the types of attributes that can be used to select an audience for employment, housing, and credit advertisements."); Galen Sherwin & Esha Bhandari, *Facebook Settles Civil Rights Cases by Making Sweeping Changes to Its Online Ad Platform*, ACLU (Mar. 19, 2019, 2:00 PM), https://www.aclu.org/blog/womens-rights/womens-rights-workplace/facebook-settles-civil-rights-cases-making-sweeping [https://perma.cc/UZ4R-9757].

[185] BENKLER ET AL, *supra* note 22, at 271–75*; see also* Kim, *supra* note 29, at 871 ("[Even if] an advertiser uses neutral targeting criteria and intends to reach a diverse audience, an ad-targeting algorithm may distribute information about opportunities in a biased way.").

[186] BENKLER ET AL., *supra* note 22, at 272–75.

[187] Cohen, *Internet Utopianism and the Practical Inevitability of Law*, *supra* note 141 and accompanying text.

as they turn advertisements into compelling, personalized narratives and *create a context of vulnerability*.[188] Intermediaries and advertisers use cognitive psychology to influence human decisions in unsuspecting ways.[189] They target the intuitive, emotional, and instinctive mode of thought ("System 1"), while bypassing the deliberative mode ("System 2").[190] To do so, they use non-informational marketing strategies.[191]

First, they stimulate users' feelings, causing emotional responses to advertisements,[192] such as sadness, happiness, fear,[193] or anxiety,[194] thereby increasing the advertisements' impact.[195]

Second, intermediaries can utilize artificial intelligence entities ("AI agents") in advertising to provide a persuasive, interactive experience by imitating human feedback.[196] AI agents are designed to engage on a social level, create a natural interactive experience between humans and algorithms, and confuse users into trusting them

[188] HARTZOG *supra* note 99, at 202 ("Precision advertising can be used to exploit biases and perpetuate falsehoods in significantly corrosive ways.").

[189] Berman, *supra* note 181, at 517–18 (referring to subconscious targeting).

[190] KAHNEMAN, *supra* note 103 and accompanying text; *see also* Shmuel I. Becher & Yuval Feldman, *Manipulating, Fast and Slow: The Law of Non-Verbal Market Manipulations*, 38 CARDOZO L. REV. 101, 112 (2016).

[191] Berman, *supra* note 181, at 522 (discussing the collapse of the informational paradigm in marketing) ("[M]arketers (1) are most successful when emotional content—not information—is presented to consumers, (2) can carefully craft marketing appeals (using humor and other non-informational techniques) to increase the viewer's/reader's receptivity to the marketing message while disengaging critical faculties, and (3) can influence consumer behavior without consumers being aware of the powerful effect of advertising."); Becher & Feldman*,* *supra* note 190, at 119–21 (referring to non-verbal market manipulation, such as the colors on shopping sites and music played in shopping centers).

[192] Tamara R. Piety, *Advertising as Experimentation on Human Subjects*, 19 ADVERT. & SOC'Y Q., no. 2, 2018, at 18 ("[M]arketers often rely on stimulating fear, anxiety, jealousy, lust, avarice, hunger, and insecurity; in short, a whole repertoire of emotions and desires.").

[193] *Id.* at 34 ("Advertising professionals readily admit that fear can sell products. Indeed, a great deal of research has been directed at attempting to find the 'optimal' level of fear. As one textbook puts it, 'the appeal to fear is especially effective as a means of enhancing motivation.'").

[194] *Id.* at 35 ("A good deal of the fear that advertising attempts to stimulate is perhaps more appropriately described as 'anxiety'—usually about conforming to social norms in dress, grooming, attractiveness, and weight.").

[195] *See* MARANTZ, *supra* note 88, at 79 ("There are as many ways to attract a person's attention as there are to bait a mousetrap, and some baits work better than others.").

[196] PASQUALE, *supra* note 33, at 89–91 (2020); ARAL, *supra* note 19, at 218–20.

as human,[197] rendering users vulnerable to manipulation.[198] Consequently, intermediaries have greater power than traditional advertisements to alter user experience and decision-making in support of a politician.

Third, intermediaries can enhance the "quality" of a message through fake "likes" and "shares." For instance, potential voters may assume that many people support a politician due to a sea of likes and re-tweets created by an army of bots.[199] They can also lead users to believe that a central entity in the social network, such as an "opinion leader," supports a politician.[200]Advertisers may create *deepfakes* that seem reliable, even though they do not reflect the truth.[201] For example, they can target deepfake videos of an opinion leader supporting a politician, even though it never happened.[202] Further, intermediaries can plant messages in the social network without disclosing they are posting on behalf of the platform or an advertiser.[203] This strategy induces subliminal trust based on false

[197] Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 993 (2019) ("[B]ots are software programs that run according to instructions. We use the term here to refer to automated agents that initiate communication online, by phone, or through other technologically mediated means."); Alexander Tsesis, *Marketplace of Ideas, Privacy, and the Digital Audiences*, 94 NOTRE DAME L. REV. 1585, 1621 (2019).

[198] WALDMAN, *supra* note 79, at 136 (2018) (referring to the false trust that social robots create. Waldman focuses on physical bots, but the insights also apply to virtual robots (bots)).

[199] It should be noted that similarly, in a commercial context, advertisers lead consumers to make a false assumption that there is a high demand for a product. *See* Arunesh Mathur et al., *Dark Patterns at Scale: Findings From a Crawl of 11K Shopping Websites*, 3 PROC. ASS'N COMPUTING MACH. ON HUM.-COMPUT. INTERACTION 81:1, 81:21 (2019), https://arxiv.org/pdf/1907.07032.pdf [https://perma.cc/8ZLR-LTPC].

[200] For discussion on the importance of the message's source, see Everett M. Rogers & David G. Cartano, *Methods of Measuring Opinion Leadership*, 26 PUB. OP. Q. 435, 435 (1962) ("Opinion leaders" are individuals who "exert an unequal amount of influence on the decisions of others.").

[201] ARAL, *supra* note 19, at 54 ("[T]hat's the future of reality distortion in a world with exponentially improving GANs technology . . . .").

[202] Chesney & Citron, *supra* note 40, at 1760 (raising the problem of deep fakes that are created by general adversarial neural networks and seem to be reliable despite not reflecting the truth); Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 MD. L. REV. 892, 894 (2019); Hasen, *supra* note 45, at 542.

[203] *See* ROBERTS, *supra* note 105, at 141 (2019) ("OnlineExperts' content moderation employees [also actually created . . .] new content, seeding sites with messages and discussion points designed to encourage customer participation and engagement, and to

assumptions.[204] As a result, information cascades occur at the social network level[205] and enhance structural vulnerabilities.[206]

Data collection and analysis, targeting tools, and vast influence strategies can subvert decision-making.[207] As targeting improves, the likelihood of mobilizing voters to favor specific politicians for the wrong reasons increases. Negative fake news advertisements concerning politicians can have severe consequences on both a politician's reputation and the public interest, infringing citizens' political security and eroding democracy.[208] Therefore, combating fake news advertisements is crucial.

Following the public's concern over fake news advertisements, Twitter CEO Jack Dorsey announced that Twitter would ban political advertisements completely.[209] Yet Twitter's ad ban has exceptions: advertisements not mentioning legislation were permitted.[210]

Unlike Twitter, Facebook did not ban political ads. Hundreds of Facebook employees objected to this policy and signed a letter to CEO Mark Zuckerberg, decrying the company's decision to allow politicians to post false claims in advertisements on the platform.[211]

bring a positive face of the brand or product. All of this activity was done surreptitiously without OnlineExperts' employees ever identifying themselves as such.").

[204] For more on this practice, see Laura E. Bladow, *Worth the Click: Why Greater FTC Enforcement Is Needed to Curtail Deceptive Practices in Influencer Marketing*, 59 WM. & MARY L. REV. 1123, 1128 (2018).

[205] For discussion of informational cascades, see Sunstein & Hastie, *supra* note 76, at 12. Informational cascades are generated when individuals follow the statements or actions of predecessors and do not express their opposing opinions because they believe their predecessors are right. *Id.* at 12–14.

[206] Susser et al., *supra* note 98, at 40 (explaining structural vulnerabilities). For information on the influences of the social network, see Jonathan Zittrain, *Engineering an Election*, 127 HARV. L. REV. F. 335, 335 (2014).

[207] ARAL, *supra* note 19, at 168 (referring to five main targeting strategies: network targeting, referral marketing, social advertising, viral design, and influencer marketing).

[208] Lavi, *supra* note 55, at 444; Anthony J. Gaughan, *Illiberal Democracy: The Toxic Mix of Fake News, Hyperpolarization, and Partisan Election Administration*, 12 DUKE J. CONST. L. & PUB. POL'Y 57, 68 (2017).

[209] It should be noted that there is no clear definition of the term "political advertisement" for these purposes. "[W]hat is or is not a political message is often in the eye of the beholder." *See* Conger, *supra* note 39.

[210] Kantrowitz, *supra* note 39.

[211] *Read the Letter Facebook Employees Sent to Mark Zuckerberg About Political Ads*, N.Y. TIMES (Oct. 28, 2019), nyti.ms/350LEW6 [https://perma.cc/6GD8-HJQL] (arguing

This letter, however, did not change Zuckerberg's decision, and he has continued to rationalize his decision on the grounds of protecting freedom of expression.[212] One week before the election, Facebook began banning new political ads from running.[213] After the election, Facebook lifted the ban on political ads and now, the site neither bans nor fact-checks political advertisements.[214] How does the law react to dissemination of fake news? The next Part provides an overview of United States law governing intermediary liability for dissemination of defamatory false content.

II. INTERMEDIARY LIABILITY: THE LAW AND NORMATIVE FREEDOM OF EXPRESSION CONSIDERATIONS

A. *American Internet Exceptionalism: The Law in the United States*

In "A Declaration of the Independence of Cyberspace," John Perry Barlow pronounced that cyberspace is not subject to traditional laws and regulations, representing a new approach known as *internet exceptionalism*.[215] Under this approach, because the internet

that free speech and paid speech are not the same thing. By allowing politicians to lie in advertisements, the platform does not protect voices. Instead, it allows politicians to use the platform as a weapon "by targeting people who believe that the content posted by political figures is trustworthy.").

[212] Josh Constine, *Zuckerberg Defends Politician Ads That Will Be 0.5% of 2020 Revenue*, TECHCRUNCH (Oct. 30, 2019, 5:32 PM), tcrn.ch/32YnHxn [https://perma.cc/P7XK-XRPL]; Isaac & Kang, *supra* note 36.

[213] Steve Kovach, *Facebook to Ban New Political Ads in Week Before Presidential Election*, CNBC, https://www.cnbc.com/2020/09/03/facebook-to-ban-political-ads-in-week-before-presidential-election.html [https://perma.cc/T5BE-XY5M] (Sept. 3, 2020, 8:24 AM).

[214] Kurt Wagner, *Facebook Still Won't Fact-Check Political Ads Headed into Election Season*, TIME (Jan. 9, 2020, 11:40 AM), https://time.com/5762234/facebook-political-ads-election/ [https://perma.cc/A3YH-SGP3]; *Facebook to End Ban on Political Ads in United States*, NBC NEWS, https://www.nbcnews.com/tech/tech-news/facebook-end-ban-political-ads-united-states-rcna336 [https://perma.cc/4CU5-KSDL] (Mar. 3, 2021, 4:44 PM).

[215] John Perry Barlow was a cyber-libertarian and digital rights activist that founded the Electronic Frontier Foundation, a nonprofit organization for preserving personal freedoms and online civil liberties. John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), projects.eff.org/~barlow/Declaration-Final.html [https://perma.cc/F7Q4-BH68]. This Article was written after the

is different from other media that preceded it, "the government should not burden it with traditional laws and regulations."[216] Internet exceptionalism is at the heart of Section 230, which directs: "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[217] Under subsection (c), titled "Protection for 'Good Samaritan' blocking and screening of offensive material,"[218] Congress declared that online intermediaries could never be treated as "publishers" of material they did not develop.[219]

In passing Section 230, Congress sought to overrule *Stratton Oakmont, Inc. v. Prodigy Services Co.*,[220] in which Prodigy's good faith efforts to monitor its site resulted in increased liability.[221] Legislators sought to promote self-regulation, free speech, and foster the rise of vibrant internet enterprises.[222] According to Section 230, intermediaries, such as Facebook and Twitter, are immune from liability for third-party content, including content provided by advertisers.

Courts have interpreted Section 230 broadly and blocked lawsuits against intermediaries.[223] This overall immunity reflects the U.S.'s strong bias toward free speech above other values and its presumption against speech restrictions.[224]

enactment of the Communications Decency Act in 1996, arguing that that the cyberspace legal order would reflect the ethical deliberation of the community instead of the coercive power that characterized real-space governance. *See* KOSSEFF, *supra* note 48, at 78.

[216] KOSSEFF, *supr*a note 48, at 78.

[217] 47 U.S.C. § 230(c)(1).

[218] 47 U.S.C. § 230(c). The "Protection for 'Good Samaritan'" subsection aims to promote self-regulation by intermediaries and encourage screening of offensive materials without bearing liability. *See* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 407 (2017).

[219] *See* 47 U.S.C. §230(c).

[220] 1995 N.Y. Misc. LEXIS 229, at *1 (N.Y. Sup. Ct. May 24, 1995).

[221] *Id.*, at *13; KOSSEFF, *supra* note 48, at 46–55.

[222] *See* Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639, 651–52 (2014).

[223] KOSSEFF, *supra* note 48, at 146; Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 NOTRE DAME L. REV. 33, 36 (2019).

[224] Oreste Pollicino & Marco Bassini, *Free Speech, Defamation and the Limits of Freedom of Expression in the EU: A Comparative Analysis*, *in* RESEARCH HANDBOOK ON

1. Failure to Censor Harmful Content

In *Zeran v. America Online, Inc.*,[225] an anonymous user advertised on an America Online, Inc. ("AOL") message board that Zeran was selling t-shirts glorifying the Oklahoma City bombing—even though he did nothing of the sort—and instructed people to call Zeran's home phone number.[226] Consequently, angry AOL subscribers rang Zeran's phone incessantly.[227] Zeran sued AOL in federal court, claiming the company negligently failed to immediately remove the false harmful post upon notification.[228] The Fourth Circuit held that distributors constituted a subset of publishers and were therefore immune from liability in accordance with Section 230.[229] Under this interpretation, Section 230 grants immunity to site hosts even if they fail to act upon knowledge of potentially illegal content on their sites.[230]

After *Zeran*, Section 230 repeatedly shielded web enterprises from lawsuits in a plethora of cases.[231] For example, in *Blumenthal v. Drudge*,[232] the court upheld immunity even when the intermediary (AOL) paid an independent contractor to write gossip columns for

EU INTERNET LAW 508, 519, 540 (Andrej Savin & Jan Trzaskowski eds., 2014) (demonstrating that in the U.S., freedom of speech protections are stronger than in the EU; the different balance courts provide between free speech and reputation is even more prominent in the digital context). For criticism, see FRANKS, *supra* note 89, at 172 (referring to the broad interpretation of online free speech as the "cult of constitution" that uses free speech to protect powerful media giants at the expense of the free speech of victims of harmful speech).

[225] Zeran v. Am. Online, Inc., 129 F.3d 327, 328–29, 332 (4th Cir. 1997); KOSSEFF, *supra* note 48, at 83.

[226] *Zeran*, 129 F.3d at 329.

[227] *Id.*

[228] *Id.* at 330.

[229] *Id.* ("By its plain language, § 230 creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service.").

[230] *Id.* at 334.

[231] *See* Lavi, *supra* note 68, at 867–70 (2016); *see generally* Caraccioli v. Facebook, Inc., 167 F. Supp. 3d 1056 (N.D. Cal. 2016) (holding that the immunity applies even when the intermediary knew of the defamatory content and did not remove it); Herrick v. Grindr, LLC, 765 F. App'x 586 (2d Cir. 2019).

[232] Blumenthal v. Drudge, 992 F. Supp. 44, 51, 53 (D.D.C. 1998); KOSSEFF, *supra* note 48, at 101.

the site that contained defamatory statements.[233] AOL was not liable even though the intermediary could exercise editorial control over its contractors.[234] Judge Friedman explained that Congress made a policy choice to provide immunity in these cases, even where the interactive service provider has an active role in making available content prepared by others.[235]

2. Editorial Decisions to Remove, or Restrict Content and Accounts

Major platforms are "crafted around two different precepts: proportionality and probability. That is, content moderation [is] a question of systemic balancing," considering the inevitability of error and choosing what kinds of errors to prefer.[236] Immunity applies when intermediaries restrict user content.[237] In contrast to Section 230(c)(2) which specifically grants immunity for online services blocking or removing third-party content,[238] Section 230(c)(1) does not subject intermediaries to a good faith obligation in doing so.[239] In fact, Section 230(c)(1) "encourages online publishers to exercise their editorial discretion, [which] ensures the publishers will 'discriminate' against some content in favor of other content."[240] Thus, courts rejected lawsuits against intermediaries that restricted user-made content.

[233] *Blumenthal*, 992 F. Supp. at 52.

[234] *Id.*

[235] *Id.*

[236] Evelyn Douek, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 763 (2021).

[237] 47 U.S.C. § 230(c)(1)–(2).

[238] Eric Goldman, *The Ten Most Important Section 230 Rulings*, 20 TUL. J. TECH. & INTELL. PROP. 1, 6–7 (2017) [hereinafter Goldman, *The Ten Most Important Section 230 Rulings*]; Eric Goldman*, Online User Account Termination and 47 U.S.C. § 230(c)(2)*, U.C. IRVINE L. REV. 659, 666 (2012) [hereinafter Goldman, *Online User Account Termination*] ("Several § 230(c)(2) cases have held that good faith is determined subjectively, not objectively. In that circumstance, courts should accept any justification for account termination proffered by the online provider, even if that justification is ultimately pretextual.").

[239] *See generally* Goldman, *The Ten Most Important Section 230 Rulings*, *supra* note 238.

[240] Eric Goldman, *Per Section 230*, *Facebook Can Tell This Plaintiff to Piss Off—Fyk v. Facebook*, TECH. & MKTG. L. BLOG (June 14, 2020), https://blog.ericgoldman.org/archives/2020/06/per-section-230-facebook-can-tell-this-plaintiff-to-piss-off-fyk-v-facebook.htm [https://perma.cc/8GUZ-8PEQ].

In *Prager University v. Google LLC*, YouTube placed videos published by Prager University in a "restricted mode," blocking third parties from advertising on videos and restricting the videos' availability.[241] In response, Prager filed an action against YouTube's editorial decision, claiming infringement of their First Amendment rights and asserting that YouTube was biased against conservative content in their video restrictions.[242]

The court in the Northern District of California rejected the claim, summarizing that decisions to restrict or make available content do not transform YouTube into a content developer.[243] Thus, the court affirmed YouTube's immunity.[244] Prager also failed to persuade the Court that YouTube's services are functionally equivalent to a traditional public forum, because platforms necessarily reflect editorial discretion rather than serving as an open "town square."[245] The Ninth Circuit upheld the dismissal of the lawsuit against Google and YouTube, concluding that YouTube was a "private forum" despite its "ubiquity" and public accessibility.[246] Thus, hosting videos did not make YouTube a "state actor" for purposes of the First Amendment.[247]

Courts have continued to find that Section 230 immunizes intermediaries for editorial decisions to moderate content. In *Fyk v. Facebook, Inc.*,[248] the Ninth Circuit concluded that blocking pictures of a man urinating was protected by the First Amendment and that "nothing in [Section] 230(c)(1) turns on the alleged motives underlying the editorial decisions."[249]

As previously mentioned, after Trump's social media use relating to the Capitol riot,[250] both Facebook and Twitter banned his

[241] Prager Univ. v. Google LLC, 2018 WL 1471939 (N.D. Cal. Mar. 26, 2018), *aff'd*, 951 F.3d 991 (9th Cir. 2020).
[242] *Id.* at *5.
[243] *Id.* at *6.
[244] *Id.* at *15.
[245] *Id.* at *5–6.
[246] Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020).
[247] *Id.*
[248] Fyk v. Facebook, Inc., 808 F. App'x 597, 598 (9th Cir. 2020); Goldman, *supra* note 240.
[249] *Fyk*, 808 F. App'x at 598.
[250] Barry & Frenkel, *supra* note 13.

accounts on their respective platforms.[251] Accordingly, Twitter and Facebook cannot bear liability for this decision.[252] Moreover, as in the case of riot incitement, where content severely violates community standards, a platform's decision to suspend an account may be a proportionate response considering the high probability that more, similarly violative posts could follow. Accordingly, Facebook's Oversight Board[253] recently upheld Trump's suspension; albeit criticism followed regarding the imposition of "the indeterminate and standardless penalty of indefinite suspension."[254]

However, when a state actor or government official moderates a public account, courts have held the First Amendment applies to protect viewpoint-based user content. In fact, government officials' accounts or platforms can be considered public forums.[255] In *Knight First Amendment Institute at Columbia University v. Trump*,[256] the Second Circuit concluded that Trump's blocking of social media users violated the users' First Amendment rights. Accordingly, by using his personal Twitter account, Trump acted in his official governmental capacity. Blocking certain users from seeing or interacting with his tweets was sufficient to establish state action and trigger First Amendment protections applicable when the government restricts speech in a public forum.[257] Trump appealed this decision to the Supreme Court. After his term expired, the Court vacated the decision and remanded it to the Second Circuit with instructions to dismiss the case because Trump was no longer President.[258]

[251] Isaac & Conger, *supra* note 14; Fung, *supra* note 16.

[252] 47 U.S.C. § 230(c)(1).

[253] For further information on the oversight board, see generally Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418 (2020).

[254] *Oversight Board Upholds Former President Trump's Suspension, Finds Facebook Failed to Impose Proper Penalty*, OVERSIGHT BD. (May 2021), https://oversightboard.com/news/226612455899839-oversight-board-upholds-former-president-trump-s-suspension-finds-facebook-failed-to-impose-proper-penalty/ [https://perma.cc/F8LR-3TMK].

[255] *See, e.g.*, Knight First Amend. Inst. v. Trump, 928 F.3d 226 (2d Cir. 2019); Davison v. Randall, 912 F.3d 666 (4th Cir. 2019); Attwood v. Clemons, 526 F. Supp. 3d 1152 (N.D. Fla. 2021).

[256] *Knight First Amend. Inst.*, 928 F.3d at 234.

[257] *Id.* at 238.

[258] Biden v. Knight First Amend. Inst., 141 S. Ct. 1220, 1221 (2021) (Thomas, J., concurring), *cert. granted* Knight First Amend. Inst. v. Trump, 928 F.3d 226 (2d Cir. 2019).

It should be noted that Justice Thomas' concurring opinion referred to the moderation power of social media platforms and criticized the discretion given to these platforms under Section 230 to screen content and block material.[259] He emphasized that "[i]t seems rather odd to say that something is a government forum when a private company has unrestricted authority to do away with it."[260] Further, he predicted that the Court will have to address the current position of the few digital platforms dominating large amounts of speech.[261] Finally, Justice Thomas stated the Court also needed to consider the ways in which legal doctrines will apply, including doctrines such as common carrier status and public accommodation to "highly concentrated, privately owned information infrastructure."[262]

In summary, platforms have editorial discretion to screen user content, suspend accounts, and block profiles. However, platforms could potentially lose complete immunity to do so if Justice Thomas' concurring opinion is adopted by the Supreme Court in the future, or if common carrier-style regulations are promulgated by Congress, narrowing platforms' discretion to moderate.[263]

3. Immunity Beyond Moderation

In *Batzel v. Smith*,[264] an operator of a website and electronic listserv for Museum Security Network ("MSN")—used for publishing posts about stolen art and other security related topics of interest to museum managers—received an email recounting a conversation in which Ellen Batzel allegedly bragged about being the

For expansion, see Eric Goldman, *Deconstructing Justice Thomas' Pro-Censorship Statement in Knight First Amendment v. Trump*, TECH. & MKTG. L. BLOG (Apr. 12, 2021), https://blog.ericgoldman.org/archives/2021/04/deconstructing-justice-thomas-pro-censorship-statement-in-knight-first-amendment-v-trump.htm [https://perma.cc/AXQ2-SSWC].

[259] *Biden*, 141 S. Ct. at 1221.

[260] *Id.*

[261] *Id.*

[262] *Id.*; Goldman, *supra* note 258.

[263] Abby Lemert & Klaudia Jaźwińska, *Justice Thomas Gives Congress Advice on Social Media Regulation*, LAWFARE (Apr. 12, 2021, 4:21 PM), https://www.lawfareblog.com/justice-thomas-gives-congress-advice-social-media-regulation [https://perma.cc/R7X5-3H99].

[264] 333 F.3d 1018 (9th Cir. 2003).

granddaughter of Heinrich Himmler (Hitler's right-hand man).[265] The person who sent the email claimed Batzel hung hundreds of old European paintings on her walls and told him she inherited them.[266] The writer believed these paintings were looted during World War II and rightfully belonged to the Jewish people.[267] Soon after receiving the email, the MSN operator made slight editorial changes and posted the defamatory and false email on their network and website, thereby making it public even though the sender did not intend to share the email.[268] Consequently, many of Batzel's clients stopped working with her.[269] Batzel sued the listserv's editor and the Netherlands Museum Association.[270] The defendants sought to dismiss the case on Section 230 grounds, but the court interpreted the term "interactive computer services" narrowly and did not to apply it to MSN, resulting in liability.[271]

On appeal, the Ninth Circuit debated the operator's responsibility for including the defamatory email in a public listserv; the court shielded it from liability, concluding that the operator should not be held responsible if a reasonable person in the same position would have believed that the sender provided the information for distribution purposes.[272] The court concluded that the listserv operator was an "interactive computer service provider" under Section 230 and immune from liability, despite its editorial control over the listserv messages.[273]

Judge Gould dissented from the majority's analysis, explaining that by providing immunity to parties that disseminate writings authors do not intend to publish, the court developed a rule that

[265] *Id.* at 1020–21.
[266] *Id.*; KOSSEFF*, supra* note 48, at 108.
[267] *Batzel*, 333 F.3d at 1021.
[268] *Id.* at 1022.
[269] *Id.*
[270] Batzel v. Smith, No. CV-00-9590, 2001 U.S. Dist. LEXIS 8929, at *4–5 (C.D. Cal. June 5, 2001).
[271] *Id.* at *21–22 ("Although several cases have held that, by virtue of the Act, internet service providers cannot be sued for defamation, none are applicable here because, unlike MSN/Cremers, the qualifying entities were true internet service providers, like America Online, that provided individuals with access to the internet."); KOSSEFF, *supra* note 48, at 109.
[272] *Batzel*, 333 F.3d at 1034.
[273] *Id.* at 1031.

encourages spreading harmful lies with impunity.[274] Judge Gould concluded that the very selection and publication of particular information on the internet forms the impression that such content is worthy of dissemination.[275]

a) Gradual Erosion of the Immunity

The first decade of Section 230's enactment represents an expansion of immunity, while the subsequent decade represents its gradual erosion.[276] First, courts determined that platforms are immune from liability only for information provided by other content providers.[277] "Information content provider" is defined as "any person or entity that is responsible, *in whole or in part*, for the *creation or development* of information provided through the internet or any other interactive computer service."[278] If a plaintiff can show that a website acted as an information content provider, then the website would receive immunity.[279] Second, Section 230 only prevents courts from treating the platform as a publisher or speaker. If a plaintiff can demonstrate that his lawsuit stemmed from an action other than publishing or speaking, a court might decide that Section 230 does not block the lawsuit.[280]

[274] *Id.* at 1038 (Gould, J., dissenting). For similar criticism of immunity for online republication, see generally Lavi, *supra* note 26, at 165; Samsel v. Desoto Cnty. Sch. Dist., 242 F. Supp. 3d 496 (N.D. Miss. 2017).

[275] *Batzel*, 333 F.3d at 1038 (Gould, J., dissenting) (explaining that the focus should not be on the author's intent, but on the defendant's actions. Thus, a defendant who has actively elected to disseminate defamatory content should not be entitled to immunity); KOSSEFF, *supra* note 48, at 113.

[276] KOSSEFF, *supra* note 48, at 166.

[277] *Id.*

[278] 47 U.S.C. § 230(f)(3) (emphasis added).

[279] *See, e.g.*, La Liberte v. Reid, 966 F.3d 79, 83–84 (2d Cir. 2020) (holding that a defendant who authored the content that accompanied the photograph of La Liberate and did not merely republish the photograph from another "information content provider" would still be liable).

[280] KOSSEFF, *supra* note 48, at 166; Agnieszka McPeak, *Platform Immunity Redefined*, 62 WM. & MARY L. REV. 1557 (2021); Gregory M. Dickinson, *Rebooting Internet Immunity*, 89 GEO. WASH. L. REV. 347 (2021) (referring to online marketplaces and arguing that "[w]here a claim is preventable other than by content moderation—for example, by redesigning an app or website—a plaintiff could freely seek relief, just as in the physical world. This approach empowers courts to identify culpable actors in the virtual world and treat like conduct alike wherever it occurs."); *see, e.g.*, HomeAway.com, Inc. v. City of Santa Monica, 918 F.3d 676, 682–84 (9th Cir. 2019) (holding that liability arose from

In 2008, a federal appellate court adopted a broader reading of the terms "responsible" and "development" under Section 230, narrowing the scope of immunity.[281] In *Fair Housing Council v. Roommates.com, LLC*,[282] a popular roommate-matching website allowed users to find roommates.[283] The website's design required users to fill out a personal profile and answer several questions, including information about gender, sexual orientation, and parental status.[284] It also required users to express their preferences with respect to roommates on each of these issues.[285]

Users selected some of the answers from drop-down menus and used an internal search engine to find roommates while filtering unfit matches according to their preferences.[286] The website also included an open-ended "additional comments" section.[287] The site periodically sent its users emails with potential roommate matches.[288] The Fair Housing Council ("FHC"), a nonprofit organization that fights housing discrimination, sued Roommates.com. The FHC alleged that the drop-down menu questions, the internal search engine, the filtering service, and the open comment section led to discrimination and violated the Fair Housing Act ("FHA").[289]

On appeal, the FHC argued that by conditioning participation in the service upon reporting restricted information, Roommates.com was an information content developer within the meaning of the statute—not a passive conduit.[290] In fact, both the website's design and

facilitating unlicensed booking transactions because a local regulation did not require the platforms to monitor or remove third-party content; it does not treat them as publishers, and thereby falls outside the preemptive scope of Section 230); Oberdorf v. Amazon.com Inc., 930 F.3d 136, 153 (3d Cir. 2019) (holding that Amazon is not immune against claims premised on other actions or failures in the sales or distribution processes), *vacated en banc*, 936 F.3d 182 (3d Cir. 2019) (certifying questions to 818 F. App'x 138 (3d Cir. 2020)); Bolger v. Amazon.com, Inc., 2020 WL 4692387 (Cal. App. Ct. Aug. 13, 2020).

[281] KOSSEFF, *supra* note 48, at 168; Lavi, *supra* note 74, at 36–41.

[282] Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008).

[283] *Id.* at 1161.

[284] *Id.*

[285] *Id.*

[286] *Id.* at 1165.

[287] *Id.* at 1162.

[288] *Id.*

[289] *Id.* at 1165, 1173.

[290] *Id.* at 1165.

questions encouraged the creation of illegal content.[291] The Ninth Circuit reversed the district court's decision, declining to grant Roommates.com immunity.[292]

Writing for the majority, Chief Justice Kozinski stressed that although the CDA established immunity, it "was not meant to create a lawless no-man's-land on the Internet."[293] By providing a limited set of prepopulated, discriminatory answers and requiring users to choose one, Roommates.com was an information content provider.[294] The site's questionnaire containing preidentified answer choices made it a developer,[295] rather than a mere "passive transmitter" of information.[296] The court also declined to grant immunity for the site's internal search engine and email mechanism because those components did not use neutral tools, but rather channeled the distribution of discriminatory content.[297] The court upheld immunity only for materials posted in the open comment section.[298] In its decision, the court referred to the "material contribution to illegality" test.[299] This test denies immunity where a defendant's own actions materially contribute to the illegality.[300] The court concluded that using *neutral tools* to carry out what may be an unlawful or illicit search does not amount to "development" for Section 230

[291] *Id.* at 1165, 1167; *see* KOSSEFF, *supra* note 48, at 170.

[292] *Roommates.com*, 521 F.3d at 1175.

[293] *Id.* at 1164; *see* KOSSEFF, *supra* note 48, at 175.

[294] *Roommates.com*, 521 F.3d. at 1165.

[295] *Id.* at 1164–65.

[296] *Id.* at 1166 ("By requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate becomes much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information.").

[297] *Id.* at 1167; *see also* Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921, 929 (9th Cir. 2007) ("By categorizing, channeling and limiting the distribution of users' profiles, Roommate provides an additional layer of information that it is 'responsible' at least 'in part' for creating or developing."), *aff'd in part en banc*, 521 F.3d 1157 (9th Cir. 2008).

[298] *Roommates.com*, 521 F.3d at 1173–74.

[299] *Id.* at 1167–68.

[300] *Id.* ("[W]e interpret the term 'development' as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus it falls within the exception to Section 230, if it contributes materially to the alleged illegality of the conduct.").

immunity.[301] In contrast, the drop-down menus led to the development of illegal, discriminatory content, and for that reason, the majority held Roommates.com liable for the discriminatory content.[302]

The dissenting opinion took a narrower view of what it means to "develop" information online.[303] Under this view, providing a drop-down menu would not constitute "creating" or "developing" information in and of itself.[304] Instead, the dissent opined that courts should examine whether the topics in drop-down menus are directly unlawful—for example, when the inquiry is a statutory violation or includes a defamatory statement.[305]

Four years later, the Ninth Circuit adopted a narrower construction, excluding roommate selection from the FHA. They reasoned that, "even though Section 230 did not protect Roommates.com from liability, the site did not commit illegal discrimination because the housing laws did not apply to roommate selection."[306] Thus, the rationale for denying immunity may no longer be applicable because discriminatory statements can be lawful in this context.[307] Yet, it is still unclear whether the previous decision barred Roommates.com from enjoying Section 230 immunity due to its general contribution to the creation of discriminatory content or because of the nature of

[301] *Id.* at 1169–72 (distinguishing between the facts of this case and other cases where intermediaries designed drop-down menus and used neutral tools)*; see also* Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1124 (9th Cir. 2003); Lindsey A. Datte, Note, *Chaperoning Love Online: Online Dating Liability and the Wavering Application of CDA § 230*, 20 CARDOZO J.L. & GENDER 769, 781 (2014); Mark D. Quist, Comment, *"Plumbing the Depths" of the CDA: Weighing the Competing Fourth and Seventh Circuit Standards of ISP Immunity Under Section 230 of the Communications Decency Act*, 20 GEO. MASON L. REV. 275, 297 (2012).

[302] *Roommates.com*, 521 F.3d at 1172.

[303] *Id.* at 1176–82 (McKeown, J., concurring in part and dissenting in part) ("The majority's unprecedented expansion of liability for Internet service providers threatens to chill the robust development of the Internet that Congress envisioned.").

[304] *Id.* at 1182.

[305] *See id.* at 1189.

[306] KOSSEFF, *supra* note 48, at 179; *see* Fair Hous. Council v. Roommates.com, LLC, 666 F.3d 1216, 1223 (9th Cir. 2012).

[307] *See Roommates.com,* 666 F.3d at 1222 ("Because we find that the FHA doesn't apply to the sharing of living units, it follows that it's not unlawful to discriminate in selecting a roommate."); Tim Iglesias, *Does Fair Housing Law Apply to "Shared Living Situations"? Or the Trouble with* Roommates, 22 J. AFFORDABLE HOUS. & CMTY. DEV. L. 111, 112 (2014).

the questions and filtering criteria themselves,[308] leaving ambiguity regarding the scope of the immunity.[309]

In *FTC v. Accusearch*, the Tenth Circuit also issued a narrow reading of Section 230 immunity.[310] Accusearch operated Abika.com, which offered customers access to private information, such as a specific cell phone's GPS location information,[311] telephone call records, and social security numbers.[312] Abika.com connected customers to third-party researchers, who obtained the desired information, and gave consumers access to the private information through Abika.com or email.[313]Abika.com publicized and promoted the purchase of details about phone calls and even offered monthly reports on call activity.[314] The Federal Trade Commission ("FTC") sued Accusearch for engaging in unfair business practices, alleging violation of Section 5(a) of the Federal Trade Commission Act.[315] The FTC claimed the site used, or caused others to use, confidential information without the data subject's authorization.[316] Accusearch argued that it was merely an interactive computer service and that the independent researchers were entirely responsible for developing the investigation reports.[317]

In a Wyoming district court, Judge Downes held that Section 230 did not immunize Accusearch because the FTC did not seek to

[308] *See* JACQUELINE D. LIPTON, RETHINKING CYBERLAW: A NEW VISION FOR INTERNET LAW 136 (2015); Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 259–60 (2018) ("We might understand the *Roommates* opinion to suggest that a provider cannot be immune when it has *knowingly* designed its service or application in order to elicit illegal third-party content. . . . As with most website developers, the company was probably very attentive to the substantive preference options from which it allowed users to choose, as well as the way it presented the choices for selection (i.e., choice architecture). But the *Roommates* court did not frame its opinion in this way.").

[309] *See generally* Varty Defterderian, Fair Housing Council v. Roommates.com*: A New Path for Section 230 Immunity*, 24 BERKELEY TECH. L.J. 563, 592 (2009); Jeff Kosseff, *The Gradual Erosion of the Law That Shaped the Internet: Section 230's Evolution Over Two Decades*, 18 COLUM. SCI. & TECH. L. REV. 1, 37 (2016).

[310] FTC v. Accusearch, 570 F.3d 1187 (10th Cir. 2009).

[311] KOSSEFF, *supra* note 48, at 181.

[312] *Accusearch*, 570 F.3d at 1191–92.

[313] *Id.* at 1190–92; KOSSEFF, *supra* note 48, at 181.

[314] KOSSEFF, *supra* note 48, at 181.

[315] 15 U.S.C. § 45(a).

[316] *Accusearch*, 570 F.3d at 1190.

[317] *Id.* at 1201.

"treat" the company as the publisher of content.[318] Moreover, even if the FTC's complaint "treated" Accusearch as a publisher, immunity would still not apply because Accusearch took part in the phone records' development by connecting users with third-party information providers and receiving an administrative fee.[319] On appeal, the Tenth Circuit affirmed Judge Downes' ruling that Section 230 did not shield Accusearch from liability under the FTC complaint.[320] The majority relied solely on Judge Downes' second line of reasoning, finding the only way Accusearch could have violated privacy laws is by publishing the private data on its website.[321] Therefore, the FTC did treat Accusearch as a publisher.[322] However, Accusearch developed the content and *made it visibly active or usable*, seeking consumer requests and coordinating with researchers.[323] The third judge denied immunity because the FTC complaint did not treat Accusearch as the publisher of the information.[324]

After *Roommates.com* and *Accusearch*, courts expressed doubts regarding internet exceptionalism and the scope of immunity,[325] leading to many contradictory judicial decisions.[326] For example, in *Dyroff v. Ultimate Software Group, Inc.*,[327] the Ninth Circuit

[318] FTC v. Accusearch, No. 06-CV-105-D, 2007 WL 4356786, at *6 (D. Wyo. Sept. 28, 2007).

[319] *Id.* ("Even if the FTC's Complaint were interpreted as 'treating' Defendants as a publisher within the meaning of the CDA, the Court believes that Defendants' claim for CDA immunity nonetheless fails to meet the requirement that the published information must have been provided by 'another information content provider.'").

[320] *Accusearch*, 570 F.3d at 1201; KOSSEFF, s*upra* note 48, at 185.

[321] *Accusearch*, 570 F.3d at 1197; *see* Accusearch 2007 WL 4356786, at *6.

[322] *Accusearch*, 570 F.3d at 1197.

[323] *Id.* at 1198.

[324] *Id.* at 1197; KOSSEFF, s*upra* note 48, at 187.

[325] *See* KOSSEFF, s*upra* note 48, at 188; Kosseff, *supra* note 309, at 22 ("My analysis demonstrates that the erosion that began with the 2008 *Roommates.com* decision has accelerated, to a point where platforms have little certainty that they will be immune from claims arising from user content.").

[326] Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093 (9th Cir. 2019) (applying immunity for tools that facilitate illegal purchases); *c.f.* Lemmon v. Snap, Inc., 995 F.3d 1085 (9th Cir. 2021) (allowing an independent negligent design claim against Snapchat to move forward because the claim did not depend on what message a Snapchat user actually sends due to a negligent design of this tool); Daniel v. Armslist, LLC, 913 N.W.2d 211, 224 (Wis. Ct. App. 2018) (denying immunity for website design features that facilitated illegal purchases), *rev'd*, 926 N.W.2d 710 (2019), *cert. denied*, 140 S. Ct. 562 (2019).

[327] *Dyroff*, 934 F.3d at 1093.

affirmed the lower court's decision[328] and upheld immunity where an intermediary engaged in data-mining and deployed machine learning algorithms, allowing it to analyze user data and channel user participation toward particular groups and specific content.[329] The court concluded that by recommending user groups and sending email notifications, Ultimate Software acted as a publisher of others' content.[330] These functions—recommendations and notifications—are tools meant to facilitate user-to-user communication and are not content in and of themselves.[331] The court concluded that the recommendation and notification functions helped facilitate this user-to-user communication, but did not materially contribute to the allegedly unlawful content.[332] The Supreme Court denied Dyroff's certiorari request.[333]

In *Daniel v. Armslist*, the website Armslist.com allowed potential buyers and sellers of firearms and ammunition to contact each other, either by clicking a link on the website or by using contact information provided by other parties.[334] This design facilitated illegal firearm purchases, one of which was used in a lethal shooting.[335] The plaintiff alleged the design and operational features of Armslist.com affirmatively "encouraged" transactions in which

[328] *Id.* Data mining and machine learning allowed the intermediary to personalize recommendations to users regarding content and discussion groups that might be of interest to the user. In some cases, the recommendations channeled users to unlawful content. In one instance, the recommendations steered a user to a discussion group dedicated to the sale of narcotics. The communication on the website allowed the user to buy heroin, who later died from consuming the heroin. *Id.* at 1094–95. The court dismissed the case ruling that recommendations to users are an ordinary, neutral function of social network websites. The intermediary used neutral tools that merely provided a framework that could be utilized for proper or improper purposes. As such, it did not "create" or "develop" the information even in part. Therefore, immunity was upheld. *Id.* at 1096–98. The situation in *Dryoff* is similar to the email service in *Roommates.com*. *Id.* at 1099; *see* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1167 (9th Cir. 2007). The court was able to reach a different conclusion because the platform gained new information from users' content and behavior in order to create a site architecture that affects behavior. *See id.* at 1165–67.

[329] *Dryoff*, 934 F.3d at 1094–95.

[330] *Id.* at 1098.

[331] *Id.*

[332] *Id.* at 1101.

[333] *See* Dryoff v. Ultimate Software Grp., Inc., 140 S. Ct. 2761 (2020).

[334] Daniel v. Armslist, LLC, 913 N.W.2d 211, 215 (Wis. Ct. App. 2018), *rev'd* 926 N.W.2d 710, 714 (Wis. 2019), *cert. denied*, 140 S. Ct. 562 (2019).

[335] *See id.* at 217.

prohibited purchasers acquired firearms.[336] The Court interpreted *Roommates.com* broadly and did not grant immunity to website design features that facilitated illegal firearm purchases, even though some sales were legal on the buyer's side.[337] However, on appeal, the Supreme Court of Wisconsin reversed the decision, reasoning that the defendant provided *neutral* tools that could be used for lawful purposes; the third parties used them to create unlawful content.[338] The court also explained that Section 230(c)(1) does not contain a good faith requirement.[339] According to the Wisconsin court, immunity applies even if the intermediary has knowledge of unlawful content on its platform and even if it designs the website to facilitate unlawful activity by omitting phone or email verification.[340]

[336] *Id.* at 215–16 (summarizing Armslist's alleged misconduct as (1) facilitating private sales by allowing users to limit searches to private sellers; (2) failing to flag "criminal" or "illegal" content; (3) warning against illegality but failing to offer specific legal guidance; (4) encouraging user anonymity; and (5) enabling buyers to evade a state waiting period that required federally-licensed firearms dealers to wait forty-eight hours after receiving a response from the background check system before transferring the firearm).

[337] Design features may thereby allow plaintiffs to bypass Section 230 and result in judicial denial of motions to dismiss, even if the design is neutral to illegality. *See id.* at 222–23; *see also* Harrington v. Airbnb, Inc., 348 F. Supp. 3d 1085, 1092 (D. Or. 2018) (requiring a user to display his picture in his profile may violate anti-discrimination law).

[338] *See Daniel*, 926 N.W.2d at 714.

[339] *Id.* at 721.

[340] "That Armslist may have known that its site could facilitate illegal gun sales does not change the result. Because § 230(c)(1) contains no good faith requirement, courts do not allow allegations of intent or knowledge to defeat a motion to dismiss." *Id.* at 726. The plaintiff filed a petition to the United States Supreme Court on this case but the petition was denied. *See* Daniel v. Armslist, LLC, 140 S. Ct. 562 (2019). *See also* Alexis Kramer, *Armslist Online Gun Sale Case Won't Get Supreme Court Review*, BLOOMBERG LAW (Nov. 25, 2019, 9:36 AM), bit.ly/2Q2BEWk [https://perma.cc/AQ5U-J3RG]. Similarly, Armslist won another ruling regarding a shooting of a police officer that was committed with a gun that was illegally purchased on Armslist*.* *See* Stokinger v. Armslist, LLC, No. 1884CV03236-F, 2020 Mass. Super. LEXIS 69, at *17 (Mass. Super. Ct. Mar. 13, 2020); *see also* Dickinson, *supra* note 280, at 391–92 (arguing that the overall immunity of Section 230 should not apply for the commercial marketplace). *But see* 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, 51 ("Section 230's liability shield has been extended to activity that has little or nothing to do with free speech, such as the sale of dangerous products. Consider Armslist.com, the self-described 'firearms marketplace.' Armslist helps match unlicensed gun sellers with buyers who cannot pass background checks . . . .").

Recently in *Lemmon v. Snap Inc.*,[341] three boys died after losing control of the wheel driving at 123 miles per hour.[342] The accident occurred after they used a Snapchat speed filter—a smartphone app designed to calculate the users' speed and show it in a photograph.[343] The parents alleged that Snapchat negligently designed unsafe products that facilitated speeding and led to the accident.[344] Based on *Fair Housing Council v. Roommates.com*,[345] the Ninth Circuit allowed the claim against Snapchat to move forward as an independent negligent design claim that does not depend on the message a Snapchat user sends.[346] This differentiated it from claims concerning content published by other content providers.[347]

As the caselaw demonstrates, courts are generally inclined to find that defendants are not information content providers—choosing to err on the side of immunity. However, some courts have challenged traditional interpretations of Section 230. Overall, the standards for excluding intermediaries from immunity remain unclear.

4. Trump's Executive Order and New Legislative Bills—an Attack on the Immunity for Moderation

The gradual erosion of immunity focused on "development of user content" and biased tools for content creation. However, recent attacks on Section 230 turned a different direction—moderation practices. After Twitter added a fact-checking label to the former President's tweets,[348] Trump attempted to curb online platforms'

[341] Lemmon v. Snap, Inc., 995 F.3d 1085, 1088 (9th Cir. 2021).

[342] *Id.*

[343] *Id.*

[344] *Id.* at 1089.

[345] 521 F.3d 1157 (9th Cir. 2008).

[346] *Lemmon*, 995 F.3d at 1093.

[347] *Id.* at 1094 ("In short, Snap 'is being sued for the predictable consequences of' designing Snapchat in such a way that it allegedly encourages dangerous behavior. *Roommates*, 521 F.3d at 1170. The CDA does not shield Snap from liability for such claims."). For further information on this ruling, see Eric Goldman, *The Ninth Circuit's Confusing Ruling Over Snapchat's Speed Filter–Lemmon v. Snap*, TECH. & MKTG. L. BLOG (May 12, 2021), https://blog.ericgoldman.org/archives/2021/05/the-ninth-circuits-confusing-ruling-over-snapchats-speed-filter-lemmon-v-snap.htm [https://perma.cc/EHX5-YLZM].

[348] *See* Makena Kelly, *Twitter Labels Trump Tweets as 'Potentially Misleading' for the First Time*, VERGE (May 26, 2020, 6:04 PM) https://www.theverge.com/2020/5/26/

protection for "good Samaritans."[349] On May 28, 2020, Trump issued the Executive Order on Preventing Online Censorship ("the Order") pertaining to online platforms.[350] Following a policy statement addressing the need to "seek transparency and accountability from online platforms, and . . . preserve the integrity and openness of American discourse and freedom of expression,"[351] the Order outlined a narrow interpretation of Section 230. It clouded the legal landscape for content moderation decisions, explaining that Section 230(c)(2) applies only to good faith moderation decisions.[352] Thus, it stripped the shield provided for moderation decisions that the government did not see as moderation in "good faith."[353] The Order further directed "all executive departments and agencies" to "ensure that their application of [S]ection 230(c) properly reflect[ed] the narrow purpose of the section and take all appropriate actions in this regard."[354]

21271207/twitter-donald-trump-fact-check-mail-in-voting-coronavirus-pandemic-california [https://perma.cc/CW2C-ENZH].

[349] Exec. Order No. 13,925, 85 Fed. Reg. 34,079, 34,080 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021) ("When an interactive computer service provider removes or restricts access to content and its actions do not meet the criteria of subparagraph (c)(2)(A), it is engaged in editorial conduct. It is the policy of the United States that such a provider should properly lose the limited liability shield of subparagraph (c)(2)(A) and be exposed to liability like any traditional editor and publisher that is not an online provider.").

[350] Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021).

[351] *Id.*

[352] *See id.* at 34,080 ("[U]nder the law, this provision is not distorted to provide liability protection for online platforms that—far from acting in 'good faith' to remove objectionable content—instead engage in deceptive or pretextual actions (often contrary to their stated terms of service) to stifle viewpoints with which they disagree. Section 230 was not intended to allow a handful of companies to grow into titans controlling vital avenues for our national discourse under the guise of promoting open forums for debate, and then to provide those behemoths blanket immunity when they use their power to censor content and silence viewpoints that they dislike. When an interactive computer service provider removes or restricts access to content and its actions do not meet the criteria of subparagraph (c)(2)(A), it is engaged in editorial conduct. It is the policy of the United States that such a provider should properly lose the limited liability shield of subparagraph (c)(2)(A) and be exposed to liability like any traditional editor and publisher that is not an online provider.").

[353] *Id.*

[354] *Id.* at 34,081.

In addition, the Order directed each executive department and agency to review media advertising expenses of online platforms and restricted platforms' receipt of advertising dollars.[355] The Department of Justice was to assess viewpoint-based speech restrictions imposed by each online platform and determine whether such platforms were problematic vehicles for government speech due to viewpoint discrimination, deception to consumers, or other bad practices.[356] The Order further provided that the White House "will submit" reports of purported "online censorship" received through its "Tech Bias Reporting Tool" to the Department of Justice and FTC.[357] The latter could "consider taking action" under applicable law, including under Section 5 of the FTC Act,[358] which makes unfair methods of competition unlawful.[359]

Legal experts agree that the Order lacked legal foundation, enforceability, and impact.[360] Recently, the Center for Democracy & Technology filed a lawsuit against it seeking invalidation.[361] In addition, the Northern District of New York ruled that the Order precluded a private right of action even if defendants arbitrarily removed a plaintiff's account or prevented him from creating a new account.[362] Recently, President Biden revoked the Order and invalidated it.[363] Therefore, it is likely that immunity provided to

[355] *See id.*

[356] *See id.*

[357] *Id.* at 34,081–82.

[358] *Id.*; *see also* 15 U.S.C. § 45.

[359] Exec. Order No. 13,925, 85 Fed. Reg. 34,079, 34,082 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021); *see also* 15 U.S.C. § 45.

[360] *See* Jan Wolfe, *Trump's Order Taking Aim at Twitter Is 'Bluster': Legal Experts*, REUTERS (May 28, 2020, 2:17 PM), reut.rs/304Bm7W [https://perma.cc/BW8F-ZY4E]; Eric Goldman, *Trump's "Preventing Online Censorship" Executive Order Is Pro-Censorship Political Theater*, TECH. & MKTG. L. BLOG (May 29, 2020), bit.ly/2B33vSk [https://perma.cc/T4N5-WMN6].

[361] *See generally* Complaint, Ctr. for Tech. and Democracy v. Trump, No. 20-1456 (D.C. Cir. June 2, 2020).

[362] *See generally* Gomez v. Zuckenburg, No. 20-633, 2020 U.S. Dist. LEXIS 130989 (N.D.N.Y. July 23, 2020). "Zuckenburg" refers to Mark Zuckerberg and is a spelling error in the original complaint. *See also* Eugene Volokh, *No Claim Against Facebook Based on President's Social Media Executive Order*, VOLOKH CONSPIRACY (July 31, 2020, 1:27 PM), bit.ly/33vRWQ8 [https://perma.cc/AZ3W-B3AA].

[363] *See* Revocation of Certain Presidential Actions and Technical Amendment, Exec. Order 14,029, 86 Fed. Reg. 27,025 (May 14, 2021); Eric Goldman & Jess Miers, *Online*

platforms under Section 230 will remain strong where platforms host and moderate third-party content.

In addition to the Order, recent legislative bills strive to narrow Section 230's immunity, attacking it from different angles and "modify[ing] the scope of protection from civil liability for 'good Samaritan' blocking and screening of offensive material."[364] Recently, a bill in Florida sought to prohibit intermediaries from deplatforming Floridian political candidates.[365] Indeed, a federal court struck the bill down.[366] Following a similar effort by Florida Governor Ron DeSantis, Texas Governor Greg Abbott signed a bill prohibiting large tech companies from blocking or restricting people and posts based on viewpoint.[367] Like the Florida law, this law will

Account Termination/Content Removals and the Benefits of Internet Services Enforcing Their House Rules, 1 J. FREE SPEECH L. 191, 193 n.5 (2021).

[364] *See* Online Freedom and Viewpoint Diversity Act, S. 4534, 116th Cong. (2020); *see also* Protecting Constitutional Rights from Online Platform Censorship Act, H.R. 83, 117th Cong. § 2 (2021) (making it unlawful for platforms to moderate "protected" content and by implication excluding illicit material from the definition of "protected"); Hannah Bloch-Wehba, *Content Moderation as Surveillance*, 36 BERKELEY TECH. L.J. (forthcoming 2022) (manuscript at n.231) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3872915) [https://perma.cc/W5L8-HMQU]) (referring to Stop the Censorship Act, H.R. 4027, 116th Cong. § 2 (2019), discussing "eliminat[ion] [of] platforms' immunity for moderating content that it deems objectionable but preserving immunity for taking down 'unlawful content' . . . ."). *But see* Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. § 5(c)(1)(A) (2020) (mandating that platforms conform with all court-ordered removal of content deemed illegal within twenty-four hours); *see generally* Kiran Jeevanjee et al., *All the Ways Congress Wants to Change Section 230*, SLATE (Mar. 23, 2021, 5:45 AM), https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html [https://perma.cc/WUB2-CQ8V].

[365] *See* Transparency in Technology Act, S.B. 7072, 2021 Leg., (Fla. 2021); *see also* Eric Goldman, *Florida Hits a New Censorial Low in Internet Regulation (Comments on SB 7072)*, TECH. & MKTG. L. BLOG (June 3, 2021), https://bit.ly/3AAgrcO [https://perma.cc/5UZT-VYMJ]. For further information, see Goldman & Miers, *supra* note 363, at 191.

[366] *See, e.g.*, NetChoice, LLC v. Moody, No. 21-220, 2021 WL 2690876, at *12 (N.D. Fla. June 30, 2021); Eric Goldman, *Florida Social Media Censorship Law ENJOINED–NetChoice v. Moody*, TECH. & MKTG. L. BLOG (June 30, 2021), https://bit.ly/3hNFsZA [https://perma.cc/RC7N-W8WP]. Florida's appeal is pending. *See* Appeal, NetChoice, No. 21cv220 (N.D. Fla. July 12, 2021).

[367] *See* H.B. 20, 87th Leg., 2d Special Sess. (Tex. 2021). For further information, see Kailyn Rhone, *Social Media Companies Can't Ban Texans Over Political Viewpoints Under New Law*, TEX. TRIB. (Sept. 2, 2021, 4:00 PM), texastribune.org/2021/09/02/Texas-social-media-censorship-legislature/ [https://perma.cc/X7QY-CNT2]; Eric Goldman, *Texas Enacts Social Media Censorship Law to Benefit Anti-Vaxxers & Spammers*, TECH. & MKTG. L. BLOG (Sept. 12, 2021), https://blog.ericgoldman.org/archives/2021/09/texas-

likely be struck down as unconstitutional.[368] However, in the shadow of potential laws, both the Order and other legislative bills might impair how intermediaries moderate content, hinder efficient moderation of harmful content (incentivizing intermediaries to act as common carriers), or chill more protected speech.[369]

B. Normative Analysis

Providing a legal structure to identify constitutional values and outlining the right balance between these values can be a difficult judgment call, albeit a crucial one. The following Part focuses on dissemination of fake news and primary situations that require nuanced examination: (1) basic intermediation; (2) moderation; (3) algorithmically personalized recommendations on organic content; and (4) targeting advertisements for profit.

Intermediary liability for defamatory fake news stories threatens freedom of speech[370] and the intermediary's freedom to conduct business using economic, technical, and financial resources.[371] However, fake news stories may also threaten public figures' reputations as members of society.[372] Furthermore, fake news stories can

enacts-social-media-censorship-law-to-benefit-anti-vaxxers-spammers.htm [https://perma.cc/38RX-MJKR].

[368] Goldman, *supra* note 367.

[369] *See* Complaint at para. 45, Ctr. for Tech. and Democracy v. Trump, No. 20-1456 (D.C. Cir. June 2, 2020) (available at https://cdt.org/wp-content/uploads/2020/06/1-2020-cv-01456-0001-COMPLAINT-against-DONALD-J-TRUMP-filed-by-CENTER-FO-et-seq.pdf [https://perma.cc/6UZF-CGBR]) ("The Order will interfere significantly with the freedom of speech of all Americans. Intermediaries that host content online will be forced to shape and apply their content moderation policies according to government officials' desires, depriving Americans of access to online forums free from government interference with their constitutionally protected speech.").

[370] The Preamble to the Constitution contains a national mandate to secure the public defense. *See* U.S. CONST. pmbl.

[371] This Charter articulates the universal values on which the EU was founded, such as dignity, solidarity, freedom, and equality. In the US, an individual's right to conduct a business or pursue an occupation is a property right. *See* Charter of Fundamental Rights of the European Union, Dec. 7, 2000, art. 16, 2000 O.J. (C 364) 1; *cf.* United States v. Arena, 180 F.3d 380, 394 (2d Cir. 1999); United States v. Santoni, 585 F.2d 667, 673 (4th Cir. 1978); Garrison v. Herbert J. Thomas Mem'l Hosp. Ass'n., 438 S.E.2d 6, 14 (W. Va. 1993).

[372] *See* Peter G. Danchin, *Defaming Muhammad: Dignity, Harm, and Incitement to Religious Hatred*, 2 DUKE F. FOR L. & SOC. CHANGE 5, 17 (2010) (referring to the values that defamation law protects).

infringe on public interest, impair public faith in the electoral system and public institutions, and harm long-term democracy.[373] How should democracies balance competing interests to protect both reputation and the public interest? In the U.S., freedom of speech enjoys stronger protections than in other Western democracies.[374] U.S. free speech jurisprudence is substantively the most speech-protective country in the world and is methodologically exceptional.[375] The purpose of this right is to shield the public from government censorship[376] and ensure the public's right to receive information.[377] Courts and scholars have developed numerous theories concerning the reason for such special free speech protections.[378] Freedom of speech promotes *individual autonomy* and *self-fulfillment*,[379] as well

[373] *See* Anthony J. Gaughan, *Illiberal Democracy: The Toxic Mix of Fake News, Hyperpolarization, and Partisan Election Administration*, 12 DUKE J. CONST. L. & PUB. POL'Y 57, 68 (2017); *see generally* Hasen, *supra* note 45; Manheim & Kaplan, *supra* note 45.

[374] *See* COHEN, *supra* note 90, at 261; Pollicino & Bassini, *supra* note 224, at 519. *But see* FRANKS, *supra* note 89, at 196–9890 (arguing that legislators, courts, and civil rights organizations have interpreted the First Amendment selectively, much like religious fundamentalists, infringing on the rights of minorities and the weak and shifting even more power from vulnerable populations to powerful ones); Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1730 (2020) ("[I]n the United States, the fundamental right of free expression protected by the First Amendment is not subject to proportionality analysis—if a court finds that there is a First Amendment right, then the First Amendment applies to the state action, and strict scrutiny normally applies.").

[375] *See* Douek, *supra* note 236, at 772 ("First, the decisionmaker asks whether or not the speech fits into a category covered by the First Amendment. Second, a series of fairly outcome-determinative rules are applied based on this categorization.").

[376] *See* NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE 10 (2015) ("The Supreme Court has interpreted the First Amendment broadly to prevent the government from censoring our speech, pushing us directly for its content, or creating legal rules that allow us to be sued for speaking the truth.").

[377] *See* Susan Nevelow Mart, *The Right to Receive Information*, 95 LAW LIBR. J. 175, 175 (2003) ("The right to receive information has evolved from its early place as a necessary corollary to the right of free speech . . . ."); *see also* Martin v. City of Struthers, 319 U.S. 141, 143 (1943).

[378] *See* RICHARDS, *supra* note 376 (reviewing influential theories that lay out justifications for the right to free speech); Balkin, *supra* note 12, at 72.

[379] *See* Joseph Raz, *Free Expression and Personal Identification*, 11 OXFORD J. LEGAL STUD. 303, 311–16 (1991) (arguing freedom of expression enables the self-determination of an individual by familiarizing the public at large with his ways of life, allowing his preferences to gain public recognition and acceptability, and reassuring that he is not alone because his experiences are known to others).

as the *search for truth*.[380] A free marketplace of ideas is essential for a liberal *democracy*.[381] Contemporary theories on democracy focus on protecting and promoting a *democratic participatory culture*.[382] Accordingly, freedom of speech is necessary to ensure an individual's ability to participate in the production and distribution of culture. This theory stresses both individual liberty and collective self-governance.[383]

The digital age and the transition from an "internet society" to an "algorithmic society" pushes freedom of expression to the forefront, raising old concerns regarding this right. The correct balance must be struck between the benefits of freedom of expression and the potential harms of fake news stories to reputation and the public interest. Intermediaries host fake news stories, providing interfaces and tools to enhance information dissemination.[384] They also use editorial discretion to decide what content to remove from the platform and what content to leave for all to see. They use algorithmic recommendations on relevant organic content that may enhance the flow of harmful content.[385] Moreover, intermediaries even target advertisements for profit by using user data and other tools and strategies to target particular advertisements to identified individuals.[386] In doing so, intermediaries allow political stakeholders to present their political messages to specific, vulnerable individuals, even though the messages may include fake news and distortions of truth. By enabling vast influence on voter consciousness, intermediaries can impair public faith in election results and erode long-term

[380] *See* JOHN STUART MILL, ON LIBERTY 5–9 (4th ed. 1869); *see generally* JOHN MILTON, AREOPAGITICA: A SPEECH FOR THE LIBERTY OF UNLICENSED PRINTING (1958).

[381] *See* ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT 83–87 (1948).

[382] *See* Jack M. Balkin, *Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 3–4 (2004).

[383] *Id.* at 3 ("Democratic culture is about individual liberty as well as collective self-governance; it is about each individual's ability to participate in the production and distribution of culture.").

[384] *See* Part III.A.

[385] *See* Part III.C.

[386] *See* Part III.D.

democracy.[387] When people spread false statements about public officials and institutions, democracy itself suffers.[388]

Arguably, the law should impose liability on intermediaries. However, imposing liability on intermediaries for fake news stories may result in collateral censorship,[389] because intermediary liability affects users' practical ability speak.[390] A traditional, individualistic understanding of speech rights does not compute new manners of free expression. Speech on social media is governed by content moderation. In such a system, allocating greater liability to digital intermediaries would cause human and algorithmic moderation to remove more legitimate content, resulting in false positives.[391] Due to liability risks, intermediaries might censor not only unprotected speech,[392] but legitimate political speech[393] that might include lies not quite reaching a level of defamation.[394] Even though scholars

[387] *See* Hasen, *supra* note 45, at 539.

[388] *See* Sunstein, *supra* note 17, at 394.

[389] *See* Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 295–96 (2011) (arguing collateral censorship occurs when a private intermediary suppresses the speech of others in order to avoid liability that otherwise might be imposed because of that speech).

[390] *See* Balkin, *supra* note 132, at 2029–32.

[391] *See* Douek, *supra* note 236, at 802 (explaining that traditional free speech theory does not fit exactly to the new system of moderation, which is not based on the individual right, but rather on proportionality and probability to err).

[392] *See* Ashcroft v. Free Speech Coal., 535 U.S. 234, 245–46 (2002) ("The freedom of speech has its limits; it does not embrace certain categories of speech, including defamation, incitement, obscenity, and pornography produced with real children."). Such categories are not entitled to freedom of expression protection because "the evil to be restricted so overwhelmingly outweighs the expressive interests, if any, at stake, that no process of case-by-case adjudication is required," and "the balance of competing interests is clearly struck." Lavi, *supra* note 105, at 530 (quoting New York v. Ferber, 458 U.S. 747, 763–64 (1982)); *see also* United States v. Alvarez, 567 U.S. 709, 717 (2012) (noting that unlike defamation, lies are protected expressions).

[393] *Cf.* Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1043–45 (2018) (explaining that legal liability and sanctions could result in censorship and consequently, legitimate speech may also be removed).

[394] The US Supreme Court struck down a portion of the Stolen Valor Act, a federal law criminalizing false statements about having a military medal, and in fact protected lies within the First Amendment. *See Alvarez*, 567 U.S. at 729–30; *see generally* Louis W. Tompros et al., *The Constitutionality of Criminalizing False Speech Made on Social Networking Sites in a Post-*Alvarez*, Social Media-Obsessed World*, 31 HARV. J.L. & TECH. 65 (2017). In addition, expressions can benefit from defamation law defenses, especially

propose to narrow First Amendment protection to exclude lies,[395] courts continue to protect fake news, as broader liability might lead to censorship of even slight inaccuracies.[396] Increased liability risks could even cause intermediaries to screen content algorithmically before it appears on the platform without transparency about the screening process, infringing speakers' autonomy, impairing the public's right to receive information, disrupting the exchange of ideas, and undermining civic and cultural participation.[397] Liability could decrease the number of relevant recommendations users receive on organic content and lead to a ban on paid political advertisements, narrowing expression opportunities for those vying for public office.[398] In addition, one might argue that by imposing liability on intermediaries, the government infringes intermediaries' rights to free speech as a speaker.

Yet, a chilling effect may be beneficial to some degree.[399] Without it, fake news stories could deplete trust and threaten the very values freedom of expression aims to protect.[400] The public's inability to distinguish truths from falsehoods could impair voters' *autonomy* to make informed choices.[401] Moreover, spreading falsehoods within online social networks could undermine truthful statements and distort competition in the *marketplace of ideas*.[402] Due to the

when they are about public figures. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).

[395] *See* Sunstein, *supra* note 17, at 421 ("The government can regulate or ban deepfakes, consistent with the First Amendment, if (1) it is not reasonably obvious or explicitly and prominently disclosed that they are deepfakes, and (2) they would create serious personal embarrassment or reputational harm.").

[396] *Id.* at 398 ("If the government is allowed to punish or censor what it characterizes as false, it might actually end up punishing or censoring truth. The reason is that its own judgments may not be reliable.").

[397] *Contra* Elkin-Koren & Perel, *supra* note 113, at 669, 672.

[398] *See* Kreiss & Perault, *supra* note 15.

[399] *See generally* SUNSTEIN, *supra* note 58.

[400] *See generally* Balkin, *supra* note 12, at 79 (explaining that when people can no longer distinguish between true and false and cease to trust others, the very same values of free speech will be impaired).

[401] *See* Gaughan, *supra* note 373, at 68.

[402] *See* FRANKS, *supra* note 89, at 119 ("[E]ven if people had strong preferences for the truth, there is no reason for confidence that the marketplace would help them discover it. As the 'fake news' epidemic has amply demonstrated."); Cass R. Sunstein, *Believing False Rumors*, *in* THE OFFENSIVE INTERNET: SPEECH, PRIVACY, AND REPUTATION 91, 102 (Saul

technological environment and intermediaries' influence on the flow of information, fake news stories can spread widely and users are more likely to perceive them as credible.[403]

However, the balance between conflicting fundamental rights should respond to intermediaries' different roles. The role an intermediary plays in the dissemination of information should affect the preemptive measures taken to combat the dissemination of fake stories.

1. Basic Intermediation

Hosting user content, designing a platform's architecture, and utilizing different communication tools all facilitate the flow of information. Basic intermediation generally enhances freedom of expression. Facilitating information dissemination, whether the content is true or false, is neutral to content but essential to a vibrant marketplace of ideas.[404] Even if a platform's interface attracts users to the service and encourages them to share more information, it typically does not aim to promote harmful speech. There are ways to nudge users into thinking reflectively before sharing harmful information; policymakers should encourage intermediaries to implement these strategies voluntarily.[405] However, holding intermediaries liable for their site's architecture and neutral communication tools is not the solution to preventing the dissemination of harmful content. Holding intermediaries liable for encouraging content sharing will disproportionately chill the flow of information. Without useful architecture and communication tools, the internet will resemble a library without a catalogue, making it difficult for users to

Levmore & Martha Craven Nusbaum eds., 2010); Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 40 (2006).

[403] Lavi, *supra* note 55, at 443 ("Within seconds, a message or a post can travel around the world and be viewed by thousands of users."); ARAL, *supra* note 19, at 28 (expanding on the rapid dissemination of lies on Twitter and how many accept them as credible).

[404] This Article focuses on general purpose platforms such as Facebook, Twitter, and YouTube. Indeed, intermediaries can operate ideological platforms for spreading political content on specific candidates and form focal points for a politician. Focal points for specific types of content are beyond the scope of this Article. *Cf.* Lavi, *supra* note 74.

[405] *Cf.* Lavi, *supra* note 55, 497–510.

find relevant information.[406] Immunity for basic intermediation is therefore necessary to promote a vibrant marketplace of ideas.

2. Moderation

Moderating users' content is one way to shape public discourse. It promotes adherence to the platforms' terms of use statements, site guidelines, and legal regimes. It is a key part of the production chain of commercial sites and social media platforms and a fundamental aspect of any platform.[407] Imposing intermediary liability for failure to remove defamatory fake news stories would force intermediaries to serve as arbiters of truth for content they neither authored nor aimed to promote. Intermediaries may remove content just because someone reported it as fake news, even if the content is not defamatory and thereby protected by the First Amendment.[408] A knowledge-based regime could result in collateral censorship, curtailing political content's availability and variety and undermining free speech. Moreover, to minimize risks, intermediaries might remove content automatically or proactively by using learning algorithms without sensitivity to context, leading to "false positives."[409] Consequently, important political criticism, satire, parody, or other statements that benefit from defamation law defenses would likely be removed.[410]

[406] *See* Seth Stern, Note, *Fair Housing and Online Free Speech Collide in* Fair Housing Council of San Fernando Valley v. Roommates.com, 58 DEPAUL L. REV. 559, 589–90 (2009) ("If all websites strictly follow the Ninth Circuit's guidance, the Internet will eventually resemble a gigantic library with no cataloging system.").

[407] *See* Part III.B.

[408] The First Amendment protects lies that do not reach the level of defamation. *See* United States v. Alvarez, 567 U.S. 709, 721 (2012).

[409] TUFECKI, *supra* note 46, at 150–51 (explaining the shortcomings of algorithmic moderation and lack of sensitivity to context); *see, e.g.*, GILLESPIE, *supra* note 27, at 98 ("These systems are just not very good yet . . . given that offense depends so critically on both interpretation and context."); NATASHA DUARTE ET AL., CTR. FOR DEMOCRACY & TECH., MIXED MESSAGES? THE LIMITS OF AUTOMATED SOCIAL MEDIA CONTENT ANALYSIS 1, 4 (2017), https://cdt.org/files/2017/11/Mixed-Messages-Paper.pdf [https://perma.cc/FKE6-4GRE].

[410] *See* DAPHNE KELLER, STAN. CTR. FOR INTERNET & SOC'Y, DOLPHINS IN THE NET: INTERNET CONTENT FILTERS AND THE ADVOCATE GENERAL'S *GLAWISCHNIG-PIESCZEK V. FACEBOOK IRELAND* OPINION 18–19 (2019) (comparing false positives to "dolphins in the net" and referring to the consequences of imposing an obligation on intermediaries to automatically screen harmful content).

Subjecting immunity to "good faith" requirements, as outlined in Trump's Order, is also undesirable.[411] It is unclear what decisions would constitute "moderation in good faith." Such scienter would undermine the motivation for "good Samaritan"[412] moderation practices and lead intermediaries to refrain from voluntary moderation in hopes of mitigating exposure to liability. Furthermore, intermediaries might not outline community guidelines to avoid viewpoint-based speech restrictions that could be perceived as "unfair" by the FTC.[413] As a result, platforms would be overused by spammers and filled with cacophony. Consequently, it would be harder for participants to find relevant content. Platforms are also likely to be abused by bad actors, filling platforms with negative content. This would make it difficult to differentiate between "wise" and "unwise" ideas and counter false statements.[414] Moreover, moderation restrictions would impair diversity among platforms and the marketplace of ideas.

Granting immunity for certain moderation roles is essential to prevent a disproportionate chilling effect on free expression and business models, mitigate cacophony and platform abuse, and promote diverse moderation practices between platforms—all leading to a robust marketplace of ideas. Thus, the law should neither require intermediaries to censor user content nor intervene with editorial discretion in content moderation practices. Instead, the law should continue to protect good Samaritan moderation practices.

[411] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079, 34,080 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021) ("[U]nder the law, this provision is not distorted to provide liability protection for online platforms that—far from acting in 'good faith' to remove objectionable content—instead engage in deceptive or pretextual actions (often contrary to their stated terms of service) to stifle viewpoints with which they disagree.").

[412] 47 U.S.C. § 230(c) (explaining current "good Samaritan" immunity).

[413] *See* 15 U.S.C. § 45.

[414] *See* ROBERTS, *supra* note 105, at 165 ("If you open a hole on the internet . . . it gets filled with shit."); *See* James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 53–54 (2015).

3. Algorithmically Personalized Recommendations

Algorithmically personalized recommendations focus user attention on relevant content and connections.[415] An algorithmic conclusion that a user advocates for a particular political party results in more content recommendations for the user's preferred political party, including related fake news stories.[416] Users are exposed to content affirming their previous dispositions and are thus more likely to reach their threshold to pass on ideas. Prioritizing content creates echo chambers and enhances polarization by reinforcing and exacerbating users' natural inclinations.[417] By recommending personalized content, the algorithm creates feedback loops that prevent equal representation of ideas.[418]

Algorithmic recommendations are never neutral; the intermediary sets the parameters for prioritization.[419] However, intermediaries can use "policy neutral" algorithms that prioritize content according to inherent characteristics, activity, and inclinations of each user without aiming to recommend unlawful content in particular.[420] In contrast, intermediaries can program a "policy directed" algorithm

[415] *See* Part I.A.3.

[416] ARAL, *supra* note 19, at 59 (referring to this feedback loop as "the Hype Loop," which "through the interplay of machine an human intelligence, controls the flow of information over the substrate; and its medium (the smartphone, at least for now), which is the primary input /output device through which we provide information to and receive information from the Hype Machine.").

[417] *See* ZUBOFF, *supra* note 90, at 466–67.

[418] MARANTZ, *supra* note 99, at 160.

[419] FRANKS, *supra* note 89, at 186 ("While algorithms are built on data, they also 'optimize' output to parameters the company chooses, crucially, under conditions also shaped by the company.").

[420] *See* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1034 (2017) (focusing on a related context of racist completion results and explaining that the autocomplete function reflects hidden biases that exist in society and "questions that large numbers of people are asking 'when they think no-one is looking.'").

without neutrality[421] or tinker with the results ex post,[422] favoring one topic over another and promoting a specific agenda. For example, they can preference content advocating a specific political candidate, prioritize fake news stories over truths, and make unlawful content more visible.[423]

Arguably, the intermediary should bear liability for recommending defamatory content and fake news stories even if their algorithm is policy neutral. This should certainly be the case where a platform's algorithm is policy directed. The intermediary has control over its algorithmic recommendations, contrasting with the relatively limited control it has in hosting users' content. Therefore, the intermediary can reduce recommendations of *unlawful* fake news stories by designing the algorithm *ex-ante*, limiting the function to avoid recommendations of specific topics or unlawful views.[424] In

[421] Lavi, *supra* note 26, at 203; Tene & Polonetsky, *supra* note 126, at 137–38 (differentiating between policy-neutral algorithms that can in some cases reflect existing, entrenched societal biases and historical inequalities and, in contrast, policy-directed algorithms that are purposefully designed to advance a predefined policy agenda); *cf.* Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 192 (Gillespie et al. eds., 2014); Waldman, *supra* note 123, at 614.

[422] In a related context, it was revealed that Google's executives and engineers tinkered with the search results without neutrality, favoring specific businesses or increasing or decreasing the visibility of specific types of content. *See* Grind et al., *supra* note 127.

[423] Differentiating policy neutral algorithms from policy directed algorithms can be challenging because algorithms are guarded trade secrets; therefore, there are legal difficulties in imposing disclosure obligations upon them. However, as Part III shall demonstrate, using impact assessment to evaluate harm caused by biased recommendations might mitigate the problem to some degree. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 142–43 (2015); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 21 (2014); Waldman, *supra* note 123, at 614.

[424] Apple's Siri is an example of such a system with limitations by design. RONALD K. L. COLLINS & DAVID M. SKOVER, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE 27 (2018) ("[S]he sidesteps medical, legal, or spiritual counsel; she eschews criminal advice; and she prefers the precise and factual to the ambiguous and evaluative."); *see also* PASQUALE, *supra* note 33, at 12 ("Regulators will need to require responsibility-by-design to complement extant models of security-by-design and privacy-by-design."). This may involve requiring certain hard-coded audit logs, or licensing practices that explicitly contemplate problematic outcomes. Such initiatives will not simply regulate robotics and AI *post hoc*, but will also influence systems development by foreclosing some design options and encouraging others. *Cf.* Jack M. Balkin, *The Three Laws of Robotics in the Age*

this way, YouTube has already restricted its system in an effort to reduce harmful recommendations.[425]

Indeed, imposing liability in these cases may result in over-censorship of legitimate recommendations for political candidates. Yet self-censored recommendations differ from external censorship of user speech. Recommendations are machine speech, directing users to content they did not specifically seek out. However, it can be argued that content prioritization and algorithmic recommendations are a key part of commercial websites' production chains. It is the intermediary's right to conduct business and design platforms as it sees fit. Imposing liability on algorithmic recommendations could undermine the intermediary's freedom of expression.[426]

It can be argued that software, algorithms, and artificial intelligence have only secondary free speech protections.[427] Even if algorithmic recommendations constitute free speech, one should differentiate between recommendations that are policy neutral and those that are policy directed. *Policy neutral* algorithmic recommendations depend on user characteristics and activities.[428] Such

of Big Data, 78 OHIO ST. L.J. 1217, 1224 (2017); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. 697, 701 (2018).

[425] *Continuing Our Work to Improve Recommendations on YouTube*, YOUTUBE (Jan. 25, 2019), https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html [https://perma.cc/GMT9-DUEN] ("[W]e'll begin reducing recommendations of borderline content and content that could misinform users in harmful ways.").

[426] *See* Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1533 (2013); *cf.* Toni M. Massaro et al., *Siri-ously 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481, 2483–84 (2017) (suggesting ways in which AI may inspire critical engagement with free speech theory and doctrine).

[427] PASQUALE, *supra* note 33, at 109 ("Free speech protections are for people, and only secondarily (if at all) for software, algorithms, and artificial intelligence."). *See also* Lawrence Lessig, *The First Amendment Does Not Protect Replicants*, *in* SOCIAL MEDIA AND DEMOCRACY (Lee Bollinger & Geoffrey Stone eds., forthcoming 2022) (manuscript at 13) (available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3922565) ("[T]he replicant targeting the ads in Facebook's algorithm would have no presumptive constitutional protection.").

[428] Tene & Polonetsky, *supra* note 126, at 138 ("'[P]olicy-neutral algorithms,' comprise[] algorithmic processes that are largely expected to provide a neutral, objective, mathematical result. What is the most profitable location for a new business? Which result do users click on when they search for the word 'Jew?' Here, users would be surprised to discover they are being presented a manicured, edited vision of the world."); *see, e.g.*, Ignacio Siles et. al*, The Mutual Domestication of Users and Algorithmic Recommendations on Netflix*, 12 COMMC'N, CULTURE & CRITIQUE 499, 508 (2019) ("Netflix makes specific

recommendations increase the magnitude ascribed to the content but do not aim to alter the proportion of unlawful and legitimate recommendations. Neutral recommendations rely on users' personal property and activities.[429] Imposing liability on intermediaries in such cases may cast too heavy a burden on the flow of information, resulting in collateral censorship of legitimate recommendations and making it more difficult for users to find relevant information.

In contrast, *policy directed algorithms* bolster the proportion of specific types of content and views.[430] Due to the centrality of social media platforms, the process underlying the marketplace of ideas may work poorly when algorithms promote a biased agenda, as the power of intermediaries creates structurally unequal access to information.[431] If the algorithm promotes fake stories, people will focus on falsehoods rather than truths, and competition among ideas will become ineffective.[432] Policy directed algorithms do not promote users' free speech and instead can inflict severe harm.[433] Liability does not give rise to concerns about collateral censorship because liability is directed at the intermediary's own recommendations.[434]

recommendations to shape these rituals, based on the technologies users employ and the content they watch when they perform the rituals.").

[429] ARAL, *supra* note 19, at 61 ("Machine intelligence ingests our thoughts, behaviors and options and, in turn, curates the stories we see in our newsfeeds, the pictures we see on Instagram, the colleagues and dated suggested to us on LinkedIn and Tinder, and the ads we are shown alongside this content.").

[430] *See, e.g.*, FRISCHMANN & SELINGER, *supra* note 104, at 117–18 (discussing the Facebook cognition experience in which the algorithm showed users only negative, or only positive stories).

[431] Helen Norton, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441, 446 (2019).

[432] Sunstein, *supra* note 402, at 92.

[433] Policy directed algorithms can promote specific harmful content, political content, or commercial content, make it more prominent, and mislead the audience regarding its importance. Facebook's advertising algorithms already use categories of targeting that can promote hate speech. *See* Kerri A. Thompson, *Commercial Clicks: Advertising Algorithms as Commercial Speech*, 21 VAND. J. ENT. & TECH. L. 1019, 1020–21 (2019) (noting that the intermediary can promote specific content without directly targeting it and without transparency that misleads the audience).

[434] An intermediary that reaps social benefits from speech has the same incentives as the original speaker and does not need the incentives that immunity provides to facilitate speech. Whenever intermediaries function as speakers, the rationale for immunity diminishes. *See* Wu, *supra* note 389, at 297 (explaining that immunity is not the appropriate response to situations in which collateral censorship is not the problem).

Some chilling effect on the intermediary's recommendations is expected, yet this is necessary to strike the right balance between the user's fundamental right to receive information and the third party's right to reputation.[435]

One can argue that algorithmic recommendations are speech. Accordingly, a balance must be struck between the intermediary's right to free speech and the rights of third parties.[436] Imposing liability for failing to reduce recommendations on defamatory fake news and harmful content can influence the public's access to political information. Due to algorithms' limitations, restrictions on specific words would cause a decrease in recommendations on political matters in general,[437] and the efficiency of the algorithm as a functional tool will decrease.[438] The costs to freedom of expression outweigh the benefits of reducing harmful recommendations by policy neutral algorithms.

[435] Selective dissemination is much like algorithmic policy directed recommendations. *Cf.* Lavi, *supra* note 26, at 182–83 (explaining that liability can be imposed on an intermediary's functions of selective dissemination).

[436] *See* Massaro et al., *supra* note 426, at 2483–84 (suggesting ways in which AI may inspire critical engagement with free speech theory and doctrine); Wu, *supra* note 426, at 1533; *see also* Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame*, 4 GEO. L. TECH. REV. 641, 641 (2020) ("[A]lthough one might wonder whether the data-driven, algorithmic activities that enable and invite such manipulation ought to count as protected speech at all, the Court's emerging jurisprudence about the baseline coverage of constitutional protection for speech seems poised to sweep many such information processing activities within the First Amendment's ambit.").

[437] Algorithms are not sensitive enough to context; therefore, trying to avoid specific recommendations would reduce effective recommendations that are important for the public's right to information. Trying to avoid recommendations of fake news is likely to result in a decline in important political speech. In another place, I proposed that intermediaries should make efforts to reduce recommendations on *unprotected speech* that constitutes incitement to terrorism. In such cases, incitement can cost human lives. Indeed, avoiding unprotected recommendations of content containing incitement may also reduce protected speech and not just incitement to terror. However, the costs to free speech in cases of incitement are lower with regard to recommendations on political speech because reducing recommendations on content that encourages violence, even if it is protected, might not be as bad as reducing recommendations on useful political content that should enjoy higher degree of First Amendment protection. *See* United States v. Alvarez, 567 U.S. 709 (2012); *cf.* Lavi, *supra* note 105.

[438] *See* Wu, *supra* note 426, at 1517–24 (differentiating between speech and functional tools).

However, recommendations that depend on intermediary preferences to promote specific types of content and agendas are a different story. Such recommendations extend far beyond functional tools. The recommendation tool itself is an expression of the intermediary's ideas[439] or advice to users.[440] Assuming recommendations should be treated as speech, intermediaries cannot have it both ways:[441] they cannot claim to be active speakers when seeking First Amendment protection and mere navigation tools when facing tort liability. By enjoying free speech rights, intermediaries undermine their Section 230 immunity and bear liability for unlawful recommendations as speakers.[442]

4. Targeting Advertisements for Profit

Targeting advertisements deliberately promotes a political agenda or advocates for a specific politician, without neutrality and without adhering to professional norms.[443] A political advertisement

[439] Yet, Wu still tends to believe that they are functional tools. *See id.* at 1525 (referring to software navigation and map programs as cases that are harder to differentiate between communication of ideas and functionality). Another approach is that algorithms represent the message of their developers and are tied to human editorial judgement. *See* Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1479 (2013). In Part III, Collins and Skover explain that the First Amendment should protect communications in all forms relevant to human utility. *See* COLLINS & SKOVER, *supra* note 424, at 42 (explaining that for constitutional purposes, what really matters is that the receiver experiences speech—including robotic speech—as meaningful and potentially useful and valuable).

[440] In fact, this machine speech repeats user speech and at times, mimics it. Thus, this repetition promotes free speech. *See* Lavi, *supra* note 26, at 179; *see also* James Grimmelmann, *Speech Engines*, 98 MINN. L. REV. 868, 895 (2014) (explaining that algorithmic communication deserves protection primarily because it provides advice to users).

[441] However, courts have reached different conclusions regarding search engines. *See* Langdon v. Google, Inc., 474 F. Supp. 2d 622, 629–31 (D. Del. 2007) (recognizing an intermediary's right to free speech in the context of page-rank and rejecting their liability for optimization); *see also* Search King, Inc. v. Google Tech., Inc., No. 02-1457, 2003 WL 21464568, at *4, (W.D. Okla. May 27, 2003); Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1193 (2008). These rulings have been criticized in literature. *See* PASQUALE, *supra* note 423, at 167; Frank Pasquale, *Reforming the Law of Reputation*, 47 LOY. U. CHI. L.J. 515, 524–27 (2015); PASQUALE, *supra* note 423; Wu, *supra* note 426, at 1496–1503, 1527 (describing the potential harm of computer-generated speech that invites regulation).

[442] *See* RICHARDS, *supra* note 376, at 87.

[443] *See* Balkin, *Keynote*, *supra* note 12.

aims to inflate the proportion of individual's holding specific views and the magnitude ascribed to such views. Targeting political advertisements influences the context of the message by controlling the target audience, the timing of the advertisement, and how it is distributed for maximum effect. Advertisers and intermediaries have more information and power than their audiences. Therefore, equal access to "wise" and "unwise" ideas will become impossible and make it difficult to counter speech with more speech.[444] Mass targeting of fake news stories would overwhelm users and disrupt their sense of reality.[445] Moreover, secret microtargeting makes it difficult to engage in "counter speech."[446]

When an advertisement includes a negative fake news story,[447] it directs false claims to specific "receptive" individuals, thereby increasing the story's believability and likelihood of further circulation.[448] Consequently, it can inflict tremendous reputational harm, distort the truth, and infringe public interest.[449] Such negative, false advertisements do not serve the values undergirding free expression. They erode trust in political discourse and in democratic participation, providing minimal benefits to the marketplace of ideas.[450] Because the intermediary collects and analyzes user data and uses special targeting tools to promote a specific agenda, fake news advertisements can disrupt the user's sense of reality and distort the marketplace of ideas.[451]

[444] *See* Dan Laidman, *When the Slander Is the Story: The Neutral Reportage Privilege in Theory and Practice*, 17 UCLA ENT. L. REV. 74, 99 (2010); Philip M. Napoli, *What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMMC'NS L.J. 55, 69 (2018); Norton, *supra* note 431, at 442.

[445] *See* Jonathan D. Varat, *Truth, Courage, and Other Human Dispositions: Reflections on Falsehoods and the First Amendment*, 71 OKLA. L. REV. 35, 48–49 (2018).

[446] Bhagwat, *supra* note 6, at 2378–79.

[447] *See* Stewart, *supra* note 12.

[448] Targeting advertisements to individuals with low thresholds for accepting ideas can start a cascade, thus other individuals will soon follow and spread the idea as well. *See* Lavi, *supra* note 55, at 454 ("The spreading and adoption of a rumor depends on encountering individuals with low thresholds who are willing to spread it further.").

[449] *See* Hasen, *supra* note 45, at 544; Sunstein, *supra* note 17, at 394 ("[I]f people spread false statements—most obviously about public officials and institutions—democracy itself will suffer.").

[450] *See* Berman, *supra* note 181, at 515; Tsesis, *supra* note 197, at 1597.

[451] *See* Varat, *supra* note 445, at 48–49.

Allowing intermediaries to micro-target fake news advertisements with impunity can lead to undesirable consequences for reputations, freedom of speech at large, and the public interest. Liability for such targeting can be justified, seeing as collateral censorship's logic does not apply to advertiser-content in the same way.[452] Unlike user-made content, which is published immediately, intermediaries solicit advertisements and determine when to target them.[453] They can fact-check and verify advertisements before targeting or require advertisers to confirm the content's validity; alternatively, they can remove advertisements upon notice.[454] This does not cause collateral censorship of advertisements because intermediaries make much of their profit from advertisements; therefore, they will still be incentivized to run advertisements even with the risk of liability.[455]

Arguably, liability infringes upon *the intermediary's free speech rights* because targeting advertisements to specific audiences at the most effective time and manner is not only a functional tool, but rather a form of commercial speech by the intermediary.[456] However, much like policy directed algorithmic recommendations, the intermediary's right to free speech undermines its immunity to civil liability.[457]

[452] *Cf.* Wu, *supra* note 389, at 330.

[453] *See, e.g.*, Joss Fong, *Facebook Showed This Ad Almost Exclusively to Women. Is That a Problem?*, VOX (July 31, 2020, 2:33 PM), https://www.vox.com/recode/2020/7/31/21349793/facebook-ad-targeting-bias-discrimination [https://perma.cc/H57U-UFX8].

[454] Jack M Balkin, *supra* note 12, at 94 (proposing that if intermediaries bear distributors' liability for advertising, such a regime will incentivize them to supervise ads more carefully).

[455] *See id.*

[456] *See* Thompson, *supra* note 433, at 1034–35.

[457] *See* Langdon v. Google, Inc., 474 F. Supp. 2d 622, 629–31 (D. Del. 2007) (recognizing an intermediary's right to free speech in the context of page-rank and rejecting their liability for optimization); RICHARDS, *supra* note 376, at 87; *see also* Search King, Inc. v. Google Tech., Inc., No. 02-1457, 2003 WL 21464568, at *4 (W.D. Okla. May 27, 2003); Bracha & Pasquale, *supra* note 441, at 1193. These rulings have been criticized in literature. *See* PASQUALE, *supra* note 423, at 167; Pasquale, *supra* note 441, at 524–27; Wu, *supra* note 426, at 1496–1503, 1527 (describing the potential harm of computer-generated speech that invites regulation); *supra* notes 437–442 and accompanying text.

C. *Reevaluating Exceptionalism in Light of Technological Developments*

Once upon a time, people thought the internet was the harbinger of "disintermediation"—a sovereign-free medium controlled from the bottom-up by users, not subject to governmental laws and regulations.[458] This perception reflects the concept of internet exceptionalism.[459] However, today's intermediaries are not mere conduits.[460] While it may seem like any internet user can publish freely and instantly online, many intermediaries actively curate the content their users post.[461] They can promote or withhold ideas, organize the flow of information, and influence social dynamics.[462] They possess an essential role in directing user attention.[463] For example, intermediaries moderate user-generated content.[464] Different intermediaries have varying attitudes towards moderation and diverse community rules.[465] Intermediaries can also use algorithms to determine what users view.[466] Moreover, through algorithmic recommendations, they can influence what is valued, posted, and shared. They collect users' information, personalize content,[467] and manipulate meanings

[458] *See* Lavi, *supra* note 74, at 11–12.

[459] Barlow started the spirit of wide-eyed techno utopianism. *See* Barlow, *supra* note 215; *see also* MARANTZ, *supra* note 88, at 68.

[460] *See* Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2297 (2014); Derek E. Bambauer, *Middlemen*, 64 FLA. L. REV. F. 64, 64–65 (2012); Sylvain, *supra* note 308, at 268 (explaining that because intermediaries structure, sort, and sometimes sell user data, they are not passive conduits).

[461] *See* MARANTZ, *supra* note 88, at 70; Bloch-Wehba, *supra* note 364, at Part II.A. (addressing the practice of intermediaries that remove content following government pressure); Klonick, *supra* note 109, at 1601.

[462] *See* Michal Lavi, *Online Intermediaries: With Power Comes Responsibility*, JOLT DIG. (May 11, 2018), https://jolt.law.harvard.edu/digest/online-intermediaries-with-power-comes-responsibility [https://perma.cc/9Y83-UJ4J].

[463] *See id.*

[464] *See* GILLESPIE, *supra* note 27, at 5–6.

[465] *See* Shannon Bond, *Critics Slam Facebook but Zuckerberg Resists Blocking Trump's Posts*, NPR (June 11, 2020, 11:58 AM), https://n.pr/37mIoqm [https://perma.cc/8865-5YHN] ("When Trump tweeted an identical message, Twitter took the novel step of hiding the tweet behind a warning label, saying it broke its rules against glorifying violence. Zuckerberg saw it differently. Even though he was personally disgusted by the president's inflammatory rhetoric, he said, the post did not break Facebook's rules against inciting violence.").

[466] *See, e.g.*, Hern, supra note 25.

[467] *Cf.* ZUBOFF, *supra* note 90, at 8–9; VAIDHYANATHAN, *supra* note 29, at 54.

in undisclosed ways and for undisclosed purposes.[468] They micro-target advertisements to specific users at the most effective times,[469] subvert user decision-making,[470] and even threaten democracy.[471]

As technology advances and algorithmic influencers become a fundamental aspect of any platform, intermediaries' duty in moderating information flows should be reconsidered.[472] Reevaluating the role of intermediaries' is particularly important, especially in light of recent attacks on Section 230.[473]

Recent scholarship acknowledges that twenty-first century intermediaries cannot be treated as mere passive conduits and that policymakers should formulate a model to understand platforms' roles and duties.[474] Different scholars have observed intermediaries' influences in different ways and have proposed various legal obligations.[475] Even though intermediaries are private entities, some scholars have proposed that since they control the information

[468] *See* COHEN, *supra* note 90, at 96.

[469] *See, e.g.*, Thompson, *supra* note 433, at 1023 (explaining that Facebook allowed advertisers to target advertisements on specific topics to hate groups); Julia Angwin et al., *Facebook Enabled Advertisers to Reach 'Jew Haters'*, PROPUBLICA (Sept. 14, 2017, 4:00 PM), https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters [https://perma.cc/U43D-EC3V].

[470] *See* GILLESPIE, *supra* note 27, at 23 ("Platforms may not shape public discourse by themselves, but they do shape the shape of public discourse. And they know it.").

[471] *See* Zittrain, *supra* note 206, at 336; *see also* Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, GUARDIAN (Mar. 17, 2018, 6:03 PM), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [https://perma.cc/V8Z8-G9KS].

[472] *See* Cohen, *Internet Utopianism and the Practical Inevitability of Law*, *supra* note 141, at 96 ("Advancing human freedom through the absence of law was never really in the cards.").

[473] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021); *see also* Transparency in Technology Act, S.B. 7072, 2021 Leg., (Fla. 2021); H.B. 20, 87th Leg., 2d Special Sess. (Tex. 2021).

[474] Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1373 (2018); Lavi, *supra* note 55, at 463.

[475] *See, e.g.*, Orit Fischman-Afori, *Online Rulers as Hybrid Bodies: The Case of Infringing Content Monitoring*, 23 J. CONST. L. 351, 407 (2021); Jack M. Balkin, *Fixing Social Media's Grand Bargain*, *in* AEGIS PAPER SERIES 2018, at 11 (Hoover Inst., Aegis Ser. Paper No. 1814, 2018) [hereinafter *Fixing Social Media's Grand Bargain*]; Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. UNIV. L. REV. (forthcoming 2021).

infrastructures that serve the public, they should be treated as public forums,[476] or at least as hybrid bodies.[477] Thus, these scholars have argued that intermediaries should be considered state actors, thereby subjecting them to the First Amendment and other public law standards.[478] Though this was rejected by the Ninth Circuit in *Prager University v. Google*,[479] such perception is reflected in Trump's Order[480] declaring that, "[i]t is the policy of the United States that large online platforms, such as Twitter and Facebook, as the critical means of promoting the free flow of speech and ideas today, should not restrict protected speech."[481] Recently, in *Biden v. Knight First Amendment Institute*'[482] Justice Thomas criticized Section 230, emphasizing that highly-concentrated, privately-owned platforms are the infrastructure for information and noting that "[i]t seems rather

[476] *See* Kyle Langvardt, *A New Deal for the Online Public Sphere*, 26 GEO. MASON L. REV. 341, 380–81 (2018) (proposing that nonstate regulators such as online platforms can be perceived as state agencies); Langvardt, *supra* note 474, at 1353 (exploring the possibility of outlining an administrative monitoring and compliance regime to ensure that the online intermediaries content moderation policies are in line with First Amendment principles); *see also* K. Sabeel Rahman, *The New Utilities: Private Power, Public Values, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1668 (2018) (proposing to apply public utilities concept on online platforms); *cf.* Bhagwat, *supra* note 6, at 2402.

[477] *See* Fischman-Afori, *supra* note 475, at 407 (proposing that online platforms should be treated as hybrid bodies and subject them to public law standards).

[478] *See* Rahman, *supra* note 476, at 1671. It should be noted that profiles of the government and government representatives are already treated as public forums. For example, the court ruled that U.S. former-President Donald Trump could not block Twitter followers due to their dissenting views because it is a violation of their First Amendment right to participate in a "designated public forum." Knight First Amend. Inst. v. Trump, 302 F. Supp. 3d 541, 549 (S.D.N.Y. 2018), *aff'd*, 928 F.3d 226 (2d Cir. 2019), *cert. granted*, *judgment vacated sub nom.* Biden v. Knight First Amend. Inst., 141 S. Ct. 1220 (2021) (vacating the Second Circuit's opinion); *see also* Davison v. Randall, 912 F.3d 666, 688 (4th Cir. 2019), *as amended* (Jan. 9, 2019).

[479] *See* Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020) ("Despite YouTube's ubiquity and its role as a public-facing platform, it remains a private forum, not a public forum subject to judicial scrutiny under the First Amendment.").

[480] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021).

[481] *Id.* It should be noted that "[t]his sentence changed in the final draft. In the prior draft, the sentence referenced the public forum doctrine." Goldman, *supra* note 360.

[482] *See* Biden v. Knight First Amend. Inst., 141 S. Ct. 1220, 1221 (2021); *see also* Goldman, *supra* note 360.

odd to say that something is a government forum when a private company has unrestricted authority to do away with it."[483]

Imposing the full spectrum of public forum obligations on intermediaries is undesirable. Functionally, it could even cause more problems. "It would do nothing to prevent third parties from using social media to manipulate end users, stoke hatred, fear, and prejudice, or spread fake news. And because social media would be required to serve as neutral public forums, they could do little to stop this."[484] Even if social media platforms ceased curating feeds, they can still collect and harvest user data directly or through third parties,[485] as the recently leaked Facebook documents demonstrate.[486] In turn, this data could be sold to third parties who could use it on their sites (or elsewhere) and influence the flow of information.[487]

A related proposal advocates for subjecting platforms to obligations not as public forums, but rather as public utilities or monopolies.[488] This position was expressed by Justice Thomas in the Supreme Court's ruling in *Biden v. Knight First Amendment Institute*,[489] analogizing private platforms to common carriers or public

[483] *Biden*, 141 S. Ct. at 1221.

[484] Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 6; *see* Langvardt, *supra* note 474, at 1367 ("[T]he more significant difficulty with applying the state action doctrine to the platforms lies in the fact that internet platforms can 'evict' unwanted speakers without involving the courts."); *see also* Balkin, *supra* note 12, at 71 ("[T]reating social media companies as state actors or as public utilities does not solve the problems of the digital public sphere."); Jack M. Balkin, *To Reform Social Media, Reform Informational Capitalism*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 107–08 (Lee Bollinger & Geoffrey R. Stone eds., forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3925143 [https://perma.cc/5SHP-YJ57] [hereinafter *To Reform Social Media*]; Citron & Franks, *supra* note 340, at 66 ("If platforms are treated as governmental actors or their services deemed public fora, then they could not act as 'Good Samaritans' to block online abuse. This result would directly contravene the will of Section 230's drafters.").

[485] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 6.

[486] *See Facebook Sold a Rival-Squashing Move as Privacy Policy, Documents Reveal*, *supra* note 154; Skelton & Goodwin, *supra* note 154 (revealing the leaked documents and explaining that Facebook planned to use its Android app to match users' location data with mobile-phone base station IDs to deliver "location-aware" products without user consent. Facebook also gave preference to deals if they shared their user data with Facebook).

[487] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 6.

[488] *See* Rahman, *supra* note 476, at 1668.

[489] *See* Biden v. Knight First Amend. Inst., 141 S. Ct. 1220, 1224 (2021); *cf.* Goldman, *supra* note 360.

accommodators. He proposed that the Supreme Court should determine how to apply such doctrines to "highly concentrated, privately owned information infrastructure[s] . . . ."[490] Due to the importance of the services social media companies offer, some states went a step further and signed bills subjecting social media platforms to "must carry" rules.[491]

However, requiring that platforms serve all customers, carry all lawful traffic, and host all content—much like phone companies that carry all calls despite their content—might not be a functional solution. Intermediaries are different than common carriers.[492] Restricting their right to exclude will result in the same problematic results as subjecting them to public forum obligations—stripping their ability to keep services safe from scammers, spammers, and other harmful posts. [493]

A third proposal is to view intermediaries as a hybrid between a conduit and a media company.[494] Intermediaries not only host content, but also use their editorial discretion to moderate content and enforce community guidelines.[495] Moreover, they connect users, prioritize and recommend relevant content to specific users, and give preference to specific items on newsfeeds, all based on relevancy and user-retention considerations.[496] Intermediaries create ecosystems of networked journalism through personalized recommendations and targeted advertisements, thereby contributing to how news

[490] *Biden*, 141 S. Ct. at 1221.

[491] *See* Transparency in Technology Act, S.B. 7072 (Fla. 2021); H.B. 20, 87th Leg., 2d Special Sess. (Tex. 2021).

[492] *See* Lavi, *supra* note 55, at 866.

[493] *See* Balkin, *To Reform Social Media*, *supra* note 484, at 108 ("[L]aws preventing social media from moderating any content would also make it useless for most people, as social media would quickly fill with pornography and spam. But the fact that content moderation is an important function of social media does not mean that government should require it.").

[494] *See* Mary Louise Kelly, *Media or Tech Company? Facebook's Profile Is Blurry*, NPR (Apr. 11, 2018, 5:59 PM), n.pr/30ULTA7 [https://perma.cc/E9PX-LHRK].

[495] GILLESPIE, *supra* note 27, at 21("[P]latforms do, and must, moderate the content and activity of users, using some logistics of detection, review, and enforcement.").

[496] *See id.* at 43 ("As soon as Facebook changed from delivering a reverse chronological list of materials that users posted on their walls to curating an algorithmically selected subset of those posts in order to generate a News Feed, it moved from delivering information to producing a media commodity out of it.").

is made.[497] They are a key pathway to news and even surpass print newspapers as information sources.[498] Arguably, as similarities between intermediaries and media companies increase, intermediaries should be subjected to the professional norms and standards applicable to traditional media.[499] Indeed, some intermediaries already apply professional standards and restrict specific types of content through their terms of services and community policies.[500] However, the law still has a role in shaping the framework[501] by outlining duties or narrowing the scope of immunity for the roles intermediaries play.[502]

A fourth proposal by Professors Balkin, Hartzog, and Richards[503] is the concept of *information fiduciaries*.[504] This approach likens intermediaries' obligations toward user information to the fiduciary duties of doctors and lawyers with their patients and clients.[505] Much like the duties of care, confidentiality, and loyalty, the law should impose special duties on intermediaries—such as

[497] *See* Erin C. Carrol, *Platforms and the Fall of the Fourth Estate: Looking Beyond the First Amendment to Protect Watchdog Journalism*, 79 MD. L. REV. 529, 556 (2020).

[498] *See* Katherine Schaeffer, *U.S. Has Changed in Key Ways in the Past Decade, from Tech Use to Demographics*, PEW RSCH. CTR. (Dec. 20, 2019), pewrsr.ch/2PSoOLs [https://perma.cc/D48P-MZYP] ("Social media is now a key pathway to news for Americans. In 2018, for the first time, social media sites surpassed print newspapers as a news source for Americans.").

[499] *See* GILLESPIE, *supra* note 27, at 43; Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 8 (explaining that social media companies should live up to certain professional standards; for example, apply codes of ethics, promote norms of civility on the platform, reduce violent and harassing content, and be transparent regarding editorial standards).

[500] *See, e.g.*, *Community Standards: Bullying and Harassment*, FACEBOOK, www.facebook.com/communitystandards/bullying [https://perma.cc/6TAH-NJBV] ("[Facebook will] remove content that's meant to degrade or shame, including, for example, claims about someone's sexual activity.").

[501] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 10 (explaining that professional norms should apply with transparency and should not be arbitrary).

[502] *See, e.g.*, Balkin, *supra* note 12, at 94 (proposing that governments might establish distributor liability for paid advertisements to incentivize intermediaries to supervise the ads they target).

[503] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 11; Richards & Hartzog, *supra* note 475, at 4.

[504] *See generally* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016).

[505] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 12.

Facebook, Google, and Twitter—in relation to their users. Intermediaries resemble fiduciaries because, much like lawyers and doctors, they receive and even actively collect personal information[506] and are trusted to treat it with care. Therefore, some have argued that the law should impose these three duties and limit how social media companies profit from their users and beneficiaries.[507] Intermediaries should neither breach user trust nor take actions that users would consider unexpected or abusive.[508] As information fiduciaries, the platforms would have a duty not to misuse user data or otherwise manipulate users.[509] Professor Balkin further proposes that "digital businesses who want the Section 230 immunity must agree to be regulated as information fiduciaries . . . [and] allow interoperability for other applications, as long as those applications also agree to act as information fiduciaries."[510] In addition, these businesses should "allow government regulators to inspect their algorithms at regular

[506] Intermediaries obtain information that their users knowingly disseminate on their platforms and actively collect incidental information on users' platform engagement that leaves digital traces*.* *See* Susser, *supra* note 98, at 30 ("[B]oth the information individuals knowingly disseminate about themselves (e.g., when they visit websites, make online purchases, and post photographs and videos on social media) and the information they unwittingly provide (e.g., when those websites record data about how long they spend browsing them, where they are when they access them, and which advertisements they click on) reveals a great deal about who each individual is, what interests them, and what they find amusing, tempting, and off-putting."); *see also* TUROW, *supra* note 91, at 34–65 (explaining that intermediaries can collect data on consumers online by tracking browsing activities, clicks, cookies, and actual purchases); ZUBOFF, *supra* note 90, at 80 ("[T]hese include websites visited, psychographics, browsing activity, and information about previous advertisements that the user has been 'shown, selected and/or made purchases after viewing.'").

[507] Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 12–13; Lavi, *supra* note 55, at 491.

[508] *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 14; Balkin, *supra* note 504, at 1229; Jack M. Balkin, *The First Amendment in the Second Gilded Age*, 66 BUFF. L. REV. 979, 1008 (2018).

[509] It should be noted that this approach strives to impose a duty on intermediaries to operate their platforms with good faith, respect for users, and non-manipulation. The information fiduciary approach raises challenges regarding feasibility, enforceability, and scope. *See* Balkin, *Fixing Social Media's Grand Bargain*, *supra* note 475, at 14. *But see* Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 498 (2019) ("This Article seeks to disrupt the emerging consensus by identifying . . . tensions and ambiguities in the theory of information fiduciaries, as well as a number of reasons to doubt the theory's capacity to resolve them satisfactorily.").

[510] Balkin, *To Reform Social Media*, *supra* note 484, at 131.

intervals for purposes of enforcing competition law, privacy, and consumer protection obligations" to ensure trustworthy and public-regarding behavior.[511]

Intermediaries' growing influence on the information flow justifies a nuanced approach that targets exceptions to exceptionalism, adapting intermediaries' immunity based on their influences on the flow of information.[512] The following Part proposes model, context-based, nuanced guidelines for intermediary immunity that refine and target exceptions while simultaneously preserving freedom of expression.

III. Contextualizing Exceptionalism and Targeting Exceptions

Online content dissemination exists in many contexts.[513] Each context facilitates distinct kinds of expressions and interactions among users. Intermediaries' roles affect three main factors that shape the context and flow of information: (1) whether the source of the message and subsequent disseminator are influential entities or opinion leaders in the social network;[514] (2) the message's context and the way it is represented;[515] and (3) the audience in a given network that forms the situation's context.[516] Arguably, these contextual factors have even more impact than the content of the message itself.[517]

The message's source, presentation, and recipients influence the magnitude and credibility ascribed to the content and the likelihood

[511] *Id.*

[512] *See* Sylvain, *supra* note 137 ("[T]hese developments undermine any notion that online intermediaries deserve immunity because they are mere conduits for, or passive publishers of, their users' expression."); Sylvain, *supra* note 308, at 220; *see also* Balkin, *supra* note 12, at 94 (proposing a careful balance of intermediary liability and intermediary immunity rules).

[513] Notably, Jaron Lanier has decried how social media giants apply context to user generated content. *See* LANIER, *supra* note 78, at 63–65 ("Speaking through social media isn't really speaking at all. Context is applied to what you say after you say it, for someone else's profit.").

[514] *See* Lavi, *supra* note 26, at 150; *see generally* GLADWELL, *supra* note 82.

[515] *See* Lavi, *supra* note 26, at 150.

[516] *See id.*

[517] *See* Lavi, *supra* note 55, at 859; Lavi, *supra* note 26, at 151.

that users will further share it.[518] Simple changes to these factors create a new context. Dissemination of user-generated content is not uniform and should be viewed contextually. Hosting and moderating content through communication tools and editorial discretion is different than recommending specific content, rendering its repetition, and placing it prominently on a user's newsfeed. Targeted advertisements that aim to influence a specific audience have even greater influence in the online environment and the way users perceive it due to the role of the intermediary in dissemination. In some contexts, intermediaries "are as much publishers as platforms, as much media as intermediary."[519] In such cases the intermediary can be perceived as the source of the message and not just a mere platform. Differentiating between various intermediary roles allows for a better understanding of internet exceptionalism's proper scope and provides a more consistent interpretation of terms like "content creation" or "development."

To develop a nuanced policy of liability that accommodates challenges in the algorithmic society, one should consider how intermediaries' different dissemination practices can influence a message's context and importance. A second factor to consider is the causal link: who is particularly responsible for taking the information out of context? In other words, the question is whether the intermediary framed the context of dissemination, or whether the contextual change was initiated primarily by user signals.[520] Taking these axes together makes it possible to outline nuanced guidelines for intermediary liability for four main roles: (1) basic intermediation; (2) moderation; (3) algorithmically personalized recommendations; and (4) targeted advertising.

A. Basic Intermediation

Hosting user content and providing tools for dissemination incentivizes users to share all types of information, whether true or false. Intermediaries harness technology and their platforms' design

[518] *See, e.g.*, BENKLER ET AL., *supra* note 22, at 270–74, 284–85 (explaining that adoption and dissemination of messages by an influential (Breitbart) was a springboard to their wide dissemination on social media).

[519] PASQUALE, *supra* note 33, at 94.

[520] *See* Lavi, *supra* note 26, at 194.

to increase the likelihood that users will reach their threshold to disseminate ideas they would not otherwise share.[521] In this capacity, intermediaries are neutral to the type and substance of content disseminated, as long as dissemination increases profits. The platforms enhance content's circulation, increase its availability, and expand the audience of recipients by designing tools that allow users to sort through vast amounts of information and share content.[522] However, they do so by using neutral tools; they neither frame specific content items nor direct audience attention to unlawful content in particular. Rather, they are conduits for good and evil.[523] If the proportion of unlawful falsehoods increases relative to true statements, it is mainly because of the network's structure and social dynamics,[524] and less attributable to the intermediary's role as host. Users generally have equal choice to publish and disseminate whatever content they prefer. Intermediary functions take content out of context only to a mild degree. Thus, the intermediary neither creates nor develops content because the context and source of the disseminated message does not go through extensive changes.

Imposing liability on intermediaries for hosting falsehoods and designing communication tools would lead them to design fewer communication tools, making it difficult for users to exchange ideas and find relevant information. Alternatively, it would lead to prescreening of content that could cause disproportionate removal of legitimate content. The result would be an imbalanced chilling effect on speech and public dialogue concerning political issues. While hosting rarely takes user content out of context, imposing liability for this role has significant social costs. Therefore, internet exceptionalism is justified for basic intermediation and intermediaries should be immunized in this capacity.

[521] *See supra* Part II.A.

[522] Lavi, *supra* note 105, at 494.

[523] *Id.* (referring to Balkin, *The First Amendment in the Second Gilded Age*, *supra* note 508, at 997 ("[B]ecause social media companies encourage as many people as possible to use their sites, the inevitable result is incivility, trolling, and abuse.")).

[524] *See* Vosoughi et al., *supra* note 19 (explaining that lies circulate faster than truths).

B. Moderation

Moderation weeds out particular types of content through enforcement of terms of services and community guidelines. The intermediary's role in moderation is to determine what types of content to filter, screen, or hide from the public. The intermediary neither frames specific content items nor directs audience attention to content. Screening of this type preserves the online environment while neither creating nor developing content.

It is impossible for the intermediary to moderate with precise accuracy. It can fail to remove harmful, defamatory content or remove too much content, including legitimate information.[525] Imposing liability for failure to remove harmful content will result in over-moderation and aggressive collateral censorship. Even if obligations to remove content depend on user-reported, defamatory fake news items, anyone could abuse this regime to remove negative information about himself, even if true. Moderation preserves the context of public discourse by enforcing community guidelines.

Intermediaries are likely to self-regulate discourse on their platforms without legal liability. This is due to the intrinsic motivation to reduce falsehoods on their platforms and market pressures from advertisers—advertisers might stop placing monetizable advertisements on a platform due to the intermediary's failure to curb harmful expressions.[526] Intermediaries are thus likely to change their moderation policies to enhance profit.[527] Both intrinsic motivation and market forces can provoke an intermediary to voluntarily curb dissemination of false and harmful content by changing the platform's design, moderation policies, or otherwise.[528]

[525] *See* Langvardt, *supra* note 474, at 1359.

[526] *See* Brett Molina, *More Companies Halt Ads on Facebook Despite New Plans to Curb Hate Speech*, USA TODAY, https://www.usatoday.com/story/tech/2020/06/26/mark-zuckerberg-facebook-update-policies-hate-speech/3265725001/ [https://perma.cc/NUL7-UMRN] (June 29, 2020, 6:10 AM).

[527] GILLESPIE, *supra* note 27, at 168 (describing pressures of users that are in fact market pressures that led Facebook to change its policy regarding pictures of breastfeeding).

[528] *Cf.* Klonick, *supra* note 109, at 1616–30 (explaining the intrinsic motivations to moderate without legal obligations to do so); Lavi, *supra* note 55, at 497–510 (proposing to use nudges to dissuade users from publishing harmful content and embedding technological features for efficient removal of content).

Trump's Order subjected moderators to state actor obligations and good faith requirements, stripping immunity for content moderation and "selective censorship."[529] However, abiding by the Order would undermine good Samaritan practices. Intermediaries would neither use their editorial discretion to moderate content, nor set community guidelines to avoid being considered discriminatory toward specific viewpoints. Without moderation, platforms would become a library without a catalogue.[530] Moreover, subjecting moderation to "good faith" and "neutrality" requirements would hinder diversity among platforms with different attitudes toward content moderation.[531] Consequently, every platform would look like the other, impairing the marketplace of ideas.

1. Transparency of Moderation Practices and Consumer Protection

Moderation helps enforce the framework of community guidelines and changes context only to a mild degree. Internet exceptionalism is justified in this role. Accordingly, failure to remove content should be immunized over content removal or discrimination by moderation. Intermediaries are likely to self-regulate the discourse on their platforms without legal liability because of intrinsic motivation and market pressures.[532]

The desirable diversity in attitudes toward moderation on different platforms would remain, allowing everyone to find a suitable forum to express opinions, enhancing the public's rights to receive information, and facilitating free expression. This is even truer in cases of vulgar expressions that do not quite reach the level of defamation and expressions that are defamatory but might benefit from the law's defenses. As private actors, intermediaries are not subject

[529] Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021).

[530] *See infra* Part III.C.

[531] *See, e.g.*, Klonick, *supra* note 109, at 1620–21; Mike Isaac & Cecilia Kang, *While Twitter Confronts Trump, Zuckerberg Keeps Facebook Out of It*, N.Y. TIMES (May 29, 2020), www.nytimes.com/2020/05/29/technology/twitter-facebook-zuckerberg-trump.html [https://perma.cc/ADH9-FNEZ] (explaining the different attitudes of Facebook and Twitter toward moderation of fake news).

[532] *See* Klonick, *supra* note 109, at 1625–30 (explaining the intrinsic motivations to moderate without legal obligations to do so).

to the First Amendment.[533] Some intermediaries can choose to use their editorial discretion and remove, hide, validate, or label users' posts, while others can allow the same post on their platform for all to see.[534] Platforms should be transparent about their moderation practices to allow users the opportunity to find the proper forum for their expressions.

Trump's Executive Order advocated for transparency.[535] In contrast to other policy statements and the Order's potential chilling effect,[536] transparency obligations do not chill speech. Intermediaries operate platforms that function as the town square[537] and provide essential public needs, such as access to information and a space to express oneself freely.[538] Due to these central functions, subjecting intermediaries to transparency obligations is desirable. Scholars have long called for greater transparency in platforms' application of community standards and content moderation decisions.[539] Some even advocate subjecting platforms to a set of norms that govern *the process of decision-making,* such as transparency, reasoning, and judicial review.[540]

Following recent public concerns, Facebook moved toward transparency and due process in moderation.[541] Facebook

[533] *See* Prager Univ. v. Google, LLC, 951 F.3d 991, 996 (9th Cir. 2020).

[534] *See* Isaac & Kang, *supra* note 531 (comparing the different attitudes of Twitter and Facebook towards moderation).

[535] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079, 34,080 (May 28, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021) ("We must seek transparency and accountability from online platforms, and encourage standards and tools to protect and preserve the integrity and openness of American discourse and freedom of expression.").

[536] *See, e.g.*, *id.* (stripping the shield provided for moderation decisions that the government does not see as moderation in "good faith.").

[537] *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017) (acknowledging access to online social media as part of the right to freedom of speech and striking down state legislation that prevented convicted criminals from accessing social media as violative of First Amendments rights).

[538] *See* Amélie Heldt & Stephan Dreyer, *Competent Third Parties and Content Moderation on Platforms: Potentials of Independent Decision-Making Bodies from a Governance Structure Perspective*, 11 J. INFO. POL'Y 266, 268 (2021).

[539] *See* Evelyn Douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. & TECH. 1, 5 (2019).

[540] *See* Fischman-Afori, *supra* note 475 and accompanying text.

[541] *See* Klonick, *supra* note 253, at 2473–74.

established an independent decision-making body to determine the type of content users would be allowed to post.[542] It also created an oversight committee ("the Board") to review appeals regarding Facebook's content takedowns that is empowered to overrule decisions made by Facebook's moderators or executives.[543] Such a body can highlight weaknesses in a platform's policy formations, provide an independent forum for discussing disputed content moderation decisions, and allow publicly available reasoning necessary for users.[544] The Board focuses only on cases with significant real-world impact, selecting cases referred to them by Facebook and users' appeals.[545] Furthermore, the Board only focuses on content moderation; decisions regarding algorithmic content management or microtargeting are beyond its jurisdiction.[546] Yet despite its limitations, the Board is a step in the right direction toward promoting transparency—setting new precedents for both user participation in a private platform's governance and users' right to due process in content moderation.[547] Transparency would allow users to contest platforms' creation of proportionality guidelines, balance values, and enhance the legitimacy of platform policies and community guidelines.[548]

[542] *See* Nick Clegg, *Welcoming the Oversight Board*, FACEBOOK (May 6, 2020), about.fb.com/news/2020/05/welcoming-the-oversight-board/ [https://perma.cc/D2LA-C6A4]; *see also* Douek, *supra* note 539, at 28–49; Kate Klonick & Thomas E. Kadri, Opinion, *How to Make Facebook's 'Supreme Court' Work*, N.Y. TIMES (Nov. 17, 2018), nyti.ms/2Ds8Ba3 [https://perma.cc/NCY4-BWZ7].

[543] *See* Douek, *supra* note 539, at 26.

[544] *See id.* at 67–68.

[545] *See id.* at 26.

[546] *See* Klonick, *supra* note 253, at 2488 ("From a regulatory perspective, many see Facebook's creation of the Board as a display of self-regulation in order to stave off actual government regulation. The Board might also be a purposeful distraction of public attention away from more critical technological concerns like algorithmic content management or microtargeting.").

[547] *Id.* at 2492.

[548] *See* Douek, *supra* note 236, at 785–89 (explaining that without transparency, platforms are not likely to apply principles of proportionality correctly and giving an example of Facebook and Twitter's attitudes regarding President Trump's Posts and Tweets and their understanding that transparency and explanations are important).

The law should ensure transparency in moderation by treating community guidelines and moderation practices[549] as consumer protection matters. Doing so would be another step in the direction toward greater transparency and due process. Thus, intermediaries should be obligated to make their moderation practices public and adhere to them. Such duties are not revolutionary; similar duties already exist in the privacy context where policies are regulated as a matter of consumer protection. Relatedly, the FTC developed privacy jurisprudence that is equivalent to common law.[550] Similar to privacy policies, the FTC should mandate transparency in moderation practices and community standards and require adherence under Section 5 of the FTC Act—prohibiting "unfair or deceptive acts or practices in or affecting commerce."[551]

Much like privacy policies, moderation practices and the boundaries of free speech should be transparent on every platform. The FTC should have the authority to investigate and bring Section 5 actions against intermediaries that fail to adhere to their declared moderation practices. A similar idea is reflected in Trump's Executive Order[552] and can be adjusted and adopted to promote transparency. Transparency in community standards and moderation practices as a matter of consumer protection will allow users to know the boundaries of free expression before participating on a platform. Users can choose the platform most appropriate for them, thus promoting freedom of expression and diversity. This solution is superior to subjecting intermediaries to public law standards, because it preserves their status as private actors and avoids subjecting them to other public law standards that would hamper diversity and undermine moderation altogether.

[549] *See id.* at 829 ("Therefore the role of public regulation can be to turn the inward-looking and unsatisfying systems of content regulation outward.").

[550] *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 586 (2014).

[551] 15 U.S.C. § 45(a)(1); *see also* Solove & Hartzog, *supra* note 550, at 599.

[552] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079, 34,082 (June 2, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021) (asking the FTC to "consider taking action," using its existing Section 5 authority to enforce deceptive and unfair trade practices, against "entities covered by Section 230 that restrict speech in ways that do not align with those entities' public representations about those practices.").

C. *Algorithmically Personalized Recommendations*

Hosting and moderating specific types of content applies equally to all users. It is different from algorithmically personalized recommendations that may select defamatory fake news stories and deliver them to specific, receptive users. Algorithmic content selection can be an act of self-expression.[553] Intermediaries that include unlawful content in their selections and personally recommend it to users can exacerbate damages inflicted by such content.[554] The selection affects the source of the message. As a result, the public might get the impression that an intermediary's choice to recommend or prioritize specific content indicates its importance. A platform's recommendation of specific content to users influences the message, makes it more visible, and creates a framing effect.[555] Thus, users are likely to pay more attention to the information and consider it more credible.[556] This is even more true where personalized recommendations deliver content to "receptive" users that are inclined toward the content, easily surpassing their threshold to share it.[557] Consequently, the proportion of specific types of content on the platform can increase. Algorithmically selected recommendations significantly take content out of context.[558] When the algorithm recommends unlawful content, it exacerbates the harm such content inflicts.[559] Intermediaries should bear responsibility for content they preference and not blame "the algorithm" for the consequences of such prioritization.[560] Arguably, internet exceptionalism should not apply to algorithmic recommendations. Immunity for algorithmic

[553] *See* Lavi, *supra* note 26, at 196 (citing Tim Wu, *Is Filtering Censorship? The Second Free Speech Tradition*, *in* CONSTITUTION 3.0: FREEDOM AND TECHNOLOGICAL CHANGE 83, 88–89 (Jeffrey Rosen & Benjamin Wittes eds., 2011)).

[554] Daphne Keller, *Amplification and Its Discontents*, KNIGHT FIRST AMEND. INST. COLUMBIA UNIV., 3–4 (June 8, 2021), https://s3.amazonaws.com/kfai-documents/documents/aa473e4dad/8.12.2021_-Keller-New-Layout.pdf [https://perma.cc/Z4GX-K366].

[555] Individuals react to a particular choice in different ways depending on how it is presented. This is the "Framing Effect." *See* KAHNEMAN, *supra* note 103, at 374–81.

[556] *See* Lavi, *supra* note 74, at 31–32.

[557] *See id.* at 16.

[558] *See* Lavi, *supra* note 26.

[559] *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008) (explaining neutral tools).

[560] PASQUALE, *supra* note 33, at 93.

recommendations would foster irresponsibility and fail to strike a proper balance between the aforementioned normative considerations.

It can be argued that immunity is not an appropriate response where collateral censorship of users' content does not occur. An intermediary that shares the same incentives as the original speaker need not be encouraged to facilitate speech, and thus, the rationale for immunity diminishes.[561] However, one can also argue discussions regarding liability for this choice architecture are moot because algorithmic recommendations are unavoidable and never neutral.[562] One content item will always be on top of the other in a user newsfeed. Even a chronological presentation of content is not a neutral tool because it prefers a time parameter over other parameters such as the frequency of interactions with users posting organic content. Furthermore, presenting all content that a user's friends share chronologically would make it difficult for users to find relevant information and impair efficiency.

However, it is not always clear whether the incentives for recommendations reflect the intermediary's incentive to "speak" or are merely an intermediation of content.[563] The algorithm can be "policy neutral" and depend on users' activities, characteristics, and biases.[564] Such neutral algorithms reinforce user inclinations without preferring one viewpoint over another.[565] When the algorithm is policy neutral and depends only on users' features, the incentives to recommend content are more similar to incentives underlying intermediation. Because the recommendations rely only on users' characteristics and activities the causal link between the intermediary's actions and potential harm weakens. Thus, extending immunity to the intermediary is justified. Imposing liability on such

[561] *See id.* at 331–33.

[562] *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 86 (2008) (suggesting that it is pointless to discuss liability for choice architecture because it is unavoidable); Lavi, *supra* note 74, at 10.

[563] Wu, *supra* note 389, at 304–08; *see* Wu, *supra* note 553, at 85; Wu, *supra* note 426, at 1521–22.

[564] Tene & Polonetsky, *supra* note 126, at 165.

[565] *See id.*

intermediation would result in censorship of useful recommendations.[566] Liability would result in tremendous social costs on freedom of expression, exceeding the benefits of liability.

Arguably, the intermediary can reduce unlawful recommendations, even if delivering recommendations according to user signals,[567] and should be accountable for failing to do so. However, in the context of defamatory fake news, which can benefit from defamation law defenses as opposed to recommendations that incite terrorism, the social costs of liability for policy neutral, algorithmic recommendations exceed the benefits.[568] Intermediaries' efforts to voluntarily reduce recommendations of falsehoods might be desirable, but doing so under the threat of liability is problematic. Collateral censorship of legitimate recommendations bearing public importance is too high of a price for society to pay in this context, particularly when the algorithm is policy neutral.

Yet the algorithm can be "policy directed" to promote the intermediary's agenda, moving beyond mere responses to user signals, thereby representing the intermediary's biases and views.[569] Much like the Facebook cognition experiment that increased the proportion of recommendations for negative content,[570] algorithmic recommendations can increase the proportion of content in favor of a specific political candidate, including negative falsehoods about

[566] *See* Stern, *supra* note 406, at 589–90 (2009) (arguing that a narrow interpretation of the term "neutral tools" will turn the internet into a "gigantic library with no cataloging system").

[567] *See* The Youtube Team, *supra* note 425 ("[W]e'll begin reducing recommendations of borderline content and content that could misinform users in harmful ways . . . .").

[568] *See generally* Lavi, *supra* note 105 (discussing the need to reduce algorithmic recommendations that incite terrorism due to the tremendous harm they inflict and the risk of violence offline that can even cost life).

[569] Tene & Polonetsky, *supra* note 126, at 137–42 (differentiating between policy neutral algorithms and policy directed algorithms); *cf.* Pelle Guldborg Hansen & Andreas Maaløe Jespersen, *Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behavior Change in Public Policy*, 4 EUR. J. RISK REGUL. 3, 6, 9 (2013) (distinguishing given contexts that accidentally influence behavior from situations involving choice architects who intentionally attempt to alter behavior by manipulating such contexts).

[570] *See* Kramer et al., *supra* note 130.

political rivals.[571] Such algorithms channel content distribution according to the intermediary's preferences.[572] In these cases, internet exceptionalism should not apply because the recommendations reflect intermediary preferences. Therefore, a causal link can be drawn between the intermediary's preferences and the recommendations. In this capacity, the intermediary's incentives are different from those of the users who publish organic content.[573] Moreover, the intermediary does not use neutral tools.[574] In fact, the intermediary provides individualized content and becomes the information's developer, at least in part; therefore, it should not be immunized.[575]

1. Differentiating Between Types of Algorithms: The Black Box Challenge

Differentiating between policy neutral and policy directed algorithms and outlining a nuanced liability regime depending on the type of algorithm may appear a just and efficient framework. Though this Article addresses nuanced liability for algorithmically personalized recommendations, it should be noted that a recently proposed regulations in the European Union on Artificial Intelligence ("AI") reflects a similar approach by differentiating between types of algorithms. It classifies AI practices to distinguished

[571] *See* Frank Swain, *How Robots Are Coming for Your Vote*, BBC (Nov. 25, 2019), https://www.bbc.com/future/article/20191108-how-robots-are-coming-for-your-vote [https://perma.cc/S4CZ-98JA].

[572] *See* Douek, *supra* note 236, at 777–78.

[573] *See* Wu, *supra* note 389, at 304–08 (explaining the divergence of incentives).

[574] *See* Douek, *supra* note 236, at 777–78.

[575] *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1167–68 (9th Cir. 2008); Kim, *supra* note 29, at 927 ("Algorithms that control the flow of information and determine who sees what are contrary to the vision of 'maximizing user control' articulated in the statute."); Sylvain, *supra* note 308, at 218; Catherine Tremble, Note, *Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine Learning Algorithms*, 86 FORDHAM L. REV. 825, 829 (2017). Even Kosseff, who advocates for broad immunity, does not discount narrowing the immunity when platforms increasingly develop more sophisticated algorithmic based technology to process user data content. *See* KOSSEFF, *supra* note 48, at 188–89 ("As platforms increasingly develop more sophisticated algorithmic based technology to process user data, it remains to be seen whether courts will conclude that they are 'responsible' for the 'development' of illegal content. For example, if a social media site allows companies to target their job advertisements to users under forty, could the site be liable for 'developing' ads that violate employment discrimination laws? . . . [S]uch liability is possible, though far from certain.").

categories[576] and bans certain uses of AI algorithms altogether.[577] In order to ban them, or otherwise impose liability, such algorithms should be recognized as AI that poses unacceptable risks in manipulating human behavior.[578] The EU proposal is vague and the categorial ban on AI uses altogether risks impairing beneficial uses. There are also obstacles in application of this approach that are relevant to our context of algorithmically personalized recommendations. Automated algorithms recommend content in the "black box."[579] In other words, they hide the values and prerogatives enacted by the encoded rules, as well as the methods and parameters for recommending content.[580] The algorithmic analysis is opaque and difficult to challenge.[581] Additionally, one must bear in mind that algorithms are guarded trade secrets; therefore, there are legal difficulties imposing disclosure obligations on their operation.[582] Without transparency and procedural fairness, plaintiffs and courts

[576] *See* Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 final (Apr. 21 2021) [hereinafter Artificial Intelligence Act] (addressing (1) unacceptable risks (Title II); (2) high risks (Title III); (3) limited risks (Title IV); (4) minimal risks (Title IX)).

[577] *See id.*; *cf.* Thomas Burri & Fredrik von Bothmer, *The New EU Legislation on Artificial Intelligence: A Primer*, at 2 (Apr. 21, 2021), https://ssrn.com/abstract=3831424 [https://perma.cc/S8XT-JJ26] ("The proposed regulation prohibits certain uses of AI. It bans the use of AI: a) to materially distort a person's behaviour; b) to exploit the vulnerabilities of a specific group of persons; c) public social scoring and d) for real time remote biometric identification in public places.").

[578] Article 5(1) of the Artificial Intelligence Act deals with prohibited AI practices such as an "AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm." Artificial Intelligence Act, *supra* note 576. For further discussion, see Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 22 COMPUT. L. REV. INT'L (forthcoming 2021), https://arxiv.org/ftp/arxiv/papers/2107/2107.03721.pdf [https://perma.cc/BJ7K-6XTQ].

[579] PASQUALE, *supra* note 423, at 8; *see* PASQUALE, *supra* note 33, at 116; Waldman, *supra* note 123, at 618–18.

[580] PASQUALE, *supra* note 423, at 8.

[581] *See* PASQUALE, *supra* note 423, at 8–9 (explaining that the judgement of software is secret and operates under laws of secrecy and technologies of obfuscation, creating a "black box" that is difficult to challenge).

[582] *See* Dennis D. Hirsch, *From Individual Control to Social Protection: New Paradigms for Privacy Law in the Age of Predictive Analytics*, 79 MD. L. REV. 439, 481 (2020); Lavi, *supra* note 26, at 203 (referring to the type of algorithm in the case of voting systems).

lack knowledge about the type of algorithm.[583] Therefore, they also lack knowledge about whether the algorithm reflects users' characteristics, activities, and biases, or rather just those of the intermediary.[584] Intermediaries should not be immunized for algorithmic decisions, but rather should be able to contest the algorithm based on societal standards of fairness and accuracy,[585] especially given the decade's worth of research on algorithmic accountability.[586] Therefore, scholars have proposed ways to audit and attribute algorithmic systems' actions to their controllers.[587]

One way to accommodate this problem is to encourage research and public review to reveal policy directed practices. Regulators can call upon or even employ independent researchers to specifically analyze digital practices and attempt to uncover biased algorithms and platforms' manipulative practices.[588] This solution has the potential to mitigate the problem. Nevertheless, independent research would reveal only some cases of biased algorithms. Consequently, the public would be left with insufficient knowledge regarding the utilization of biased algorithms and intermediaries' manipulative influences.

Another path to accommodate this problem is process-based. Scholars have proposed a range of mechanisms, such as promoting algorithmic transparency, due process, and accountability

[583] *See* Hirsch, *supra* note 582, at 458–59; Lavi, *supra* note 26, at 203.

[584] *See* Hirsch, *supra* note 582, at 458–59; Lavi, *supra* note 26, at 203.

[585] *See* PASQUALE, *supra* note 33, at 107 ("[A]lgorithmic arrangements of information should be subject to contestation based of societal standards of fairness and accuracy. The alternative is to privilege rapid and automatic machine communication over human valued, democratic will formation, and due process.").

[586] *See* Frank A. Pasquale, *Data-Informed Duties in AI Development*, 119 COLUM. L. REV. 1917, 1937 (2019).

[587] *See* PASQUALE, *supra* note 33, at 92.

[588] *See* Lavi, *supra* note 26, at 204; *cf.* Ryan Calo & Alex Rosenblat, *The Taking Economy: Uber, Information, and Power*, 117 COLUM. L. REV. 1623, 1684 (2017).

obligations.[589] For example, a whistleblower mechanism[590] could be adopted to protect media giants' individual employees who might come forward to address issues of the flawed practices of biased personalized recommendations, thereby promoting disclosure. Other scholars have argued that the way to achieve transparency is through data protection legislation.[591] Legal protections for automated decision-making[592] and individuals' rights to receive explanations concerning algorithmic models,[593] (such as the protections included in the European Union's General Data Protection Regulation ("GDPR")),[594] are likely to achieve more transparency and procedural justice. Yet, the GDPR focuses on protection of the data subject's rights[595] and is therefore less suitable to reduce the harm algorithmic recommendations inflict on third parties.

Another idea is pre-implementation of a licensing regime. Accordingly, regulators would require companies to disclose algorithms' parameters and the methodology they employed, create an "audit trail that records the basis of the predictive decisions, both in

[589] *See* Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1308, 1314 (2020); Citron & Pasquale, *supra* note 423, at 18–27; Waldman, *supra* note 123 at 618–19 (reviewing different approaches for algorithmic transparency); *see generally* PASQUALE, *supra* note 423; Tal Z. Zarsky, *Transparent Predictions*, 2013 UNIV. ILL. L. REV. 1503 (2013).

[590] *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 126 (2019).

[591] Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 198–99 (2019).

[592] *See* Council Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR]; *see* Kaminski, *supra* note 591, at 196–98 (referring to the rights outlined by the GDPR to explanation, namely the rights to information about individual decisions made by algorithms).

[593] *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1087 (2018).

[594] *See* GDPR, *supra* note 592, at arts. 5–6 (referring to lawfulness of processing information); *id.* at arts. 13–14 (obligations of data controllers to provide information to data subjects regarding the purpose of processing their data); *id.* at art. 15 (the right of data subjects to access the data collected on them); *id.* at art. 7122 (the right of data subjects not to be subject to a decision based solely on automated processing).

[595] *Id.* at art. 1 (referring to the objectives of the GDPR), art. 2 (referring to the material scope of the GDPR).

terms of the data used and the algorithm employed[,]"[596] and give individuals and regulators alike the opportunity to access audit trails on demand.[597] This could allow regulators—such as the FTC or an agency like the Food and Drug Administration—to review algorithmic systems and protect against unfairness.[598] This approach removes the burden from individuals and places it on companies and licensors. But in doing so, it creates a regulatory bottleneck for companies that must move quickly to compete.[599] Furthermore, it involves substantial administrative costs that might not be feasible and may hinder innovation.[600]

Transparency, in the form of source code publication or an explanation of the results, sheds some light on the opaque process. However, such transparency is functionally useless to most individuals without specialized knowledge or factual evidence to determine whether the algorithm complies with the law.[601] Furthermore, the focus on documentation and process elevates a structure that promotes compliance with the law, obscures the fact that algorithmic decision-making erodes substantive values of fairness, equality, and dignity, and discourages both users and policymakers from taking more robust actions.[602]

Many scholars advocate a different solution that extends beyond the design stage—an algorithmic impact assessment.[603]

[596] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 127–28 (2014).

[597] Citron & Pasquale, *supra* note 423, at 28.

[598] *See id.*; *see also* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 115–16 (2017).

[599] *See* Hirsch, *supra* note 582, at 477.

[600] *See* Adam Thierer et al., *Artificial Intelligence and Public Policy*, MERCATUS CTR. GEO. MASON UNIV. at 18–20, 35 (2017), https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf [https://perma.cc/9KJ8-2UFF] (arguing that this solution might hinder innovation and that the creation of a new regulatory body to audit algorithms, datasets, and techniques advances a "transparency paradox" of its own).

[601] *See* Waldman, *supra* note 123, at 628–29.

[602] *See id.*

[603] *See id.* at 618, 628–29 ("Algorithmic impact assessments can identify and evaluate risks, consider alternatives, identify strategies to mitigate risks, and help articulate the rationale for the automated system."); *see also* Katyal, *supra* note 590, at 126 ("[R]egulatory monitors could question platform coders or ask to see the internal reports that those coders produced. In the context of privacy, for instance, regulators could ask platforms to provide privacy impact assessments."); Deirdre K. Mulligan & Kenneth A.

Accordingly, intermediaries would have to ensure their algorithms and tools undergo *regular* safety evaluations by independent auditors and technology experts.[604] Algorithmic impact assessments can mitigate the risk of error and failure at the design stage and decrease unexpected, unlawful recommendations.[605] This idea is not so revolutionary. Recently, legislators proposed to apply impact assessments in the discrimination context. The Algorithmic Accountability Act of 2019[606] requires entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments. Such *regular* evaluations mitigate discrimination and correct accordingly in a timely manner.[607] The need to evaluate algorithms is also reflected in proposed regulation in the EU[608] referring to high-risk AI algorithms and proposing a comprehensive risk management system when AI algorithms are brought into circulation.[609]

Indeed, such solutions are not optimal.[610] They leave opacity regarding the algorithms' functions. Further, implementation guidelines and enforcement should be outlined more clearly,[611] as any regulatory system would have to develop substantive standards.

Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 781, 830–31 (2019); Hartzog & Richards, *supra* note 374, 1758–59 (discussing the Algorithmic Accountability Act of 2019, requiring algorithmic impact assessments for "high-risk automated decision systems" to "regularly evaluate their tools for accuracy, fairness, bias, and discrimination."); Rory Van Loo, *The Missing Regulatory State: Monitoring Businesses in an Age of Surveillance*, 72 VAND. L. REV. 1563, 1575 (2019); Frank Pasquale, *The Second Wave of Algorithmic Accountability*, LAW & POL. ECON. PROJECT (Nov. 25, 2019), https://lpeproject.org/blog/the-second-wave-of-algorithmic-accountability/ [https://perma.cc/Z3MJ-Y76H].

[604] *See* Waldman, *supra* note 123, at 628–29.

[605] Lavi, *supra* note 105, at 565.

[606] *See* Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (2019). For further analysis and criticism, see Margot E. Kaminski & Andrew D. Selbst, Opinion, *The Legislation that Targets the Racist Impacts of Tech*, N.Y. TIMES (May 7, 2019), nyti.ms/2Ybb8MT [https://perma.cc/WMD7-463D].

[607] *See* H.R. 2231.

[608] *See* Artificial Intelligence Act, *supra* note 576.

[609] *See id.* at art. 9; *cf.* Burri & von Bothmer, *supra* note 577, at 4.

[610] *See* Kaminski & Selbst, *supra* note 606 (analyzing the flaws of the Algorithmic Accountability Act of 2019).

[611] *See* Hartzog & Richards, *supra* note 374, at 1759.

This solution is, however, flexible and preferable in that it provides full disclosure to the regulator.[612]

A recent proposal in this direction tasks the FTC with evaluating algorithmic impact and enforcing against unfair and deceptive algorithmic practices.[613] In order to evaluate and police harmful algorithmic practices, scholars have proposed that the FTC use its authority to prohibit "unfair or deceptive" trade practices under Section 5 of the FTC Act[614] to curb the use of algorithms that harm victims' reputations and to protect public interest. Section 5 is the most obvious existing mechanism that can regulate algorithmic dark patterns and other forms of manipulation.[615] Similar to the proposal subjecting moderation practices to the FTC's authority as a matter of consumer protection,[616] here, the FTC would create a framework and evaluate companies' algorithmic determinations to decide if they are unfair. Over time, the FTC could formulate "unfairness" precedent to which companies could refer before deploying their algorithms.[617] The "unfairness" approach would avoid a regulatory bottleneck of pre-deployment licensing requirements and task a regulator—one with a long record of substantial expertise in information economy—with evaluation.[618] Such evaluations of unfair algorithmic practices can shed light on opaque algorithmic practices.

[612] *See generally id.*

[613] *See generally* Hirsch, *supra* note at 582, at 494.

[614] Federal Trade Commission Act § 5, 15 U.S.C. § 45; *see* CHRIS J. HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY 31–53 (2016); Hirsch, *supra* note 582, at 447 ("Section 5 of the FTC Act gives the FTC the power to declare and enforce against 'unfair or deceptive [business] acts or practices.' This short phrase gives the FTC two distinct powers: the authority to enforce against business practices that are 'deceptive;' and the ability to enforce against business practices that are 'unfair.'").

[615] *See generally, e.g.*, FTC v. AMG Cap. Mgmt., 910 F.3d 417 (9th Cir. 2018), *rev'd*, 141 S. Ct. 1341 (2021); Stipulated Order for Permanent Injunction & Monetary Judgement, FTC v. Off. Depot, No. 9-19-cv-80431 (S.D. Fla. Mar. 28, 2019) (perceiving dark patterns as lies and misrepresentations); *see* Luguri & Strahilevitz, *supra* note 156, at 82–84 (explaining that legal commentators have largely failed to notice that the FTC is beginning to combat dark patterns with some success, at least in court, although not using the terminology of dark patterns).

[616] *See* Part III.B.1.

[617] *See* Hirsch, *supra* note 582, at 503–04.

[618] *See id.*

2. Liability for Algorithmic Recommendations: Remedies and Enforcement

The FTC's evaluation and enforcement of algorithmic practices would make it possible to differentiate between policy neutral and policy directed algorithms. If evaluation reveals an algorithm as policy directed, the intermediary would not be immunized. Instead, the FTC might issue a complaint that could lead to an administrative proceeding resulting in a cease and desist order. An intermediary's failure to comply and fix the biased policy directed algorithm would lead to civil penalty.[619] Such a failure could render the intermediary's practice "unfair."

Declaring a practice unfair might even pave the path for private litigation. However, defamation lawsuits might be futile. Indeed, the intermediary develops content by changing the context and creates personalized recommendations. Further, its incentives are not those of mere intermediation. However, holding the intermediary responsible *under defamation law* as direct publishers of such "machine speech" might be far-reaching.[620] Moreover, even if the law recognizes policy directed algorithmic recommendations as the intermediaries' direct publication, entities are only compensated when the falsehood reaches the level of defamation,[621] an especially difficult task for plaintiffs that are public figures or officials.[622] However, an

[619] *See* HOOFNAGLE, *supra* note 614, at 109; Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 806 (2020) ("The FTC requires companies operating under consent decrees to submit assessments roughly every two years for the life of the order. Assessments have to be completed by a 'qualified, objective, independent third-party' auditor with sufficient experience. And they must describe specific privacy controls, evaluate their adequacy given the size and scope of the company, explain how they meet FTC requirements, and certify they are operating effectively.").

[620] For further information on machine speech, see Grimmelmann, *supra* note 440.

[621] *See* United States v. Alvarez, 567 U.S. 709, 717–18 (2012) (explaining that lies are protected expressions). For criticism, see SUNSTEIN, supra note 59, at 48 ("the plurality in *Alvarez* was myopic in focusing largely on established categories of cases, such as defamation, in which false statements of fact can sometimes be regulated or sanctioned. In the modern era, false statements falling short of libel are causing serious problems for individuals and society; if they cause such problems, there is a legitimate argument that they should be regulable.").

[622] A public figure must prove the standard of actual malice, in which case defamation law defenses are narrower. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964). *Contra* Sunstein, *supra* note 17, at 413.

FTC declaration concerning unfair practices can pave the way for litigation on the grounds that a company was negligent for failing to exercise reasonable standards of care in designing algorithms that result in harm.[623]

D. Targeting of Advertisements

Targeting advertisements aims to promote a specific agenda, not just enhance engagement on the platform. The intermediary utilizes personal user data and targets advertisements for profit, using policy directed algorithms that aim to create behavioral changes that promote specific brands or agendas.[624] Moreover, the intermediary shapes the context of the information flow by using special strategies of refined ad targeting.[625] By designing targeting tools and using enormous amounts of user data, the intermediary offers advertisers the opportunity to display "the right ad, to the right person, at the right time," and manipulate users.[626]

Data-driven targeting tools and strategies of influence frame the advertisement, enhance the magnitude ascribed to it, and influence the context of the message that the advertisement promotes. The

[623] For more information regarding holding intermediaries responsible for the design of their platform, see Sylvain, *supra* note 308. Such an idea can be adopted regarding the design of algorithmic recommendations. The law has yet to develop in the field of duty of care for algorithmic design and might need to develop standards of reasonableness of algorithms. *See* RYAN ABBOT, THE REASONABLE ROBOT: ARTIFICIAL INTELLIGENCE AND THE LAW 58 (2020); Alina Glaubitz, How Should Liability Be Attributed for Harms Caused by Biases in Artificial Intelligence? 3 (senior thesis, Yale University) (Apr. 29, 2021), available at https://politicalscience.yale.edu/sites/default/files/glaubitz_alina.pdf [https://perma.cc/V5EZ-G2E5].

[624] PARISER, *supra* note 142, at 15; Thompson, *supra* note 433, at 1026 ("[T]he advertising algorithms offer the same speech over and over, limiting the marketplace of ideas to one familiar store. This kind of personalized advertising 'serve[s] up a kind of invisible auto propaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious to the dangers lurking in the dark territory of the unknown.'").

[625] Such refined personalized ad targeting shapes what one sees on social media and when he sees it and "what one sees and likes on social media may shape what one thinks and believes." *See* Thompson, *supra* note 433, at 1027 n.51; *see also* Kim, *supra* note 29, at 892 ("The platforms themselves play an important role in how job opportunities are distributed because they design the targeting or matching algorithms that control information flows.").

[626] *See* Kim, *supra* note 29, at 878.

intermediary functions as a social actor and users may even perceive it as the source of the advertisement. Due to the intermediary's centrality, users are likely to ascribe importance to the message since it originates from an influential entity.[627] By targeting vulnerable "receptive" target audiences at the right time, the intermediary also increases the likelihood that the target user will spread the advertisement.

Targeting fake news advertisements exacerbates the gravity of harm such advertisements inflict on reputation, foundationally threatening markets, integrity of elections, and democracy itself.[628] Intermediaries that use data-driven targeting participate together with the advertisers to develop the information in the advertisements.[629] The tools and strategies intermediaries use in targeting, channel the distribution of advertisements without neutrality and subvert the target's reflective decision-making.[630] Targeting does not enable user-informed choices. It exposes each user to different recommendations in light of algorithmic conclusions based on parameters set by the intermediary. This distances the user from positive and meaningful choices because targeting influences how users perceive their available choice sets.[631] Given this substantial influence narrowly targeted advertising has on context, the immunity regime can no longer be justified. Intermediaries are "responsible," at least "in part," for creating or developing illegal content, because they use data-driven information and targeting tools and co-develop content with users.[632]

[627] *See* WALDMAN, *supra* note 79, at 146; BENKLER ET AL., *supra* note 22 and accompanying text; GLADWELL, *supra* note 82 and accompanying text.

[628] *See* Jim Balsillie, *Data Is Not the New Oil – It's the New Plutonium*, FIN. POST (May 28, 2019), https://financialpost.com/technology/jim-balsillie-data-is-not-the-new-oil-its-the-new-plutonium [https://perma.cc/QU3V-DAC4].

[629] *See* Thompson, *supra* note 433, at 1023; *see generally* Balkin, *supra* note 12, at 84.

[630] *See* Fair Hous. Council v. Roommates.Com, LLC, 521 F.3d 1157, 1175 (9th Cir. 2008) (noting the differentiation between neutral tools and tools that are not neutral).

[631] *See* Kim, *supra* note 29, at 894.

[632] *See* Sylvain, *supra* note 308, at 211.

1. Liability for Targeting of Advertisements: Remedies and Enforcement

An intermediary should not be immunized for data-driven, targeted advertising for profit. Liability as the speaker of an unlawful message is over-broad since an intermediary targets a tremendous number of advertisements. The transaction between an intermediary and an advertiser is conducted by automatic auction, where algorithms make bids and target advertisements.[633] Arguably, the intermediary can fact-check political ads before running them and stop micro-targeted falsehoods.[634] Yet fact-checking every political advertisement and filtering advertisements before targeting users has its costs. It can impair the efficiency of markets and innovation. Moreover, liability can hinder free speech. Subjecting intermediaries to the same obligations as publishers would turn intermediaries into arbiters of truth. Intermediaries would over-censor advertisements because they lack tools to differentiate between true and false. In the context of advertisements, there is an extensive grey area between false or misleading content and mere puffery that is not entirely false,[635] making fake news ad-filtering an even more difficult task.

Knowledge-based distributor liability for false advertisements mitigates concerns regarding the burdening enforcement costs and over-censorship concerns. Under this regime, victims, the public, and civil society organizations would report fake news advertisements to the intermediary, and the intermediary would not bear liability if it removed the advertisements. Failure to remove an advertisement would not result in automatic liability and would only be imposed if content was false.

Indeed, reports about unlawful advertisements can be incorrect and directed at legitimate advertisements. Yet, knowledge-based distributor liability gives the intermediary the opportunity to decide

[633] *See, e.g.*, *About Smart Bidding*, GOOGLE ADS, https://support.google.com/google-ads/answer/7065882?hl=en [https://perma.cc/H3KM-4G2E].

[634] *See* Cohen, *supra* note 24.

[635] *See, e.g.*, Carlucci v. Han, 886 F. Supp. 2d 497, 522 (E.D. Va. 2012) (excluding puffery statements from legal liability); Adi Osovsky, *Puffery on the Market: A Behavioral Economic Analysis of the Puffery Defense in the Securities Arena*, 6 HARV. BUS. L. REV. 333, 336–37 (2016).

how to best handle advertising that reflects the core of its business. Knowledge-based distributor liability balances freedom of expression, dignity, and the public interest. This regime is less likely to lead to collateral censorship because intermediaries solicit advertisements for fees and would still have incentives to run ads, even if immunity is narrowed.[636] Knowledge-based distributor liability preserves the incentive to remove advertisements that include absolute falsehoods. For example, under this regime, Facebook is more likely to remove advertisements that include false conspiracies upon knowledge to avoid risking liability.[637]

Enforcement of knowledge-based distributor liability for advertisements raises difficulties. Arguably, targeting is personalized; different individuals are exposed to different advertisements, making enforcement difficult. Yet to some extent, FTC enforcement can bridge the gap, mitigate reputational harm to public figures, and protect the public interest. The FTC already addresses certain aspects of advertisements (such as disclosure requirements for endorsements)[638] and enforces violations of disclosure obligations under Section 5.[639] The FTC has imposed a legal duty on businesses to not engage in deceptive advertising and can police such practices.[640] Section 5 is vague and open to interpretation; its definitions are general and the confines of misleading and "unfair" practices remain

[636] *See generally* Balkin, *supra* note 12, at 94.

[637] *See* Stewart, *supra* note 11 (explaining the fake news story on Joe Biden that opened this Article).

[638] In its March 2013 guide, the FTC addresses how businesses can modify their practices to comport with fair advertising. While .com disclosures focus on all advertising mediums, it provides specific recommendations regarding disclosures for advertisements on social media platforms. The 2013 guide does not have the force and effect of law. Yet, non-compliance may lead to FTC enforcement actions for unfair or deceptive practices in violation of the FTC Act. There is an underlying legal duty for businesses not to engage in deceptive advertising and the guidelines articulate rules of conduct. *See generally* FED. TRADE COMM'N, .COM DISCLOSURES: HOW TO MAKE EFFECTIVE DISCLOSURES IN DIGITAL ADVERTISING (2013).

[639] Federal Trade Commission Act § 5, 15 U.S.C. § 45(a).

[640] *See id.*; COHEN, *supra* note 90, at 56 ("In the absence of a regulatory framework specifically tailored to the problems of surreptitious tracking and 'behavioral advertising,' the FTC attempted to fill the regulatory gap by asserting its general authority to police unfair and deceptive practices in commerce."); *see generally* HOOFNAGLE, *supra* note 614.

open.[641] Regardless, the FTC's baseline rule is clear: "do not lie."[642] Arguably, the FTC can initiate an interrogation or respond to user complaints regarding misleading political advertisements even though the advertisements do not market tangible products. This is because intermediaries are compensated for targeting users and are likely to mislead consumers of internet services.

The FTC has broad investigatory authority that provides the basis for enforcement.[643] The FTC does not generally monitor platforms.[644] It starts investigations in response to complaints by the Consumer Sentinel Network,[645] political candidates, civil society organizations, and members of Congress.[646] It resolves pending investigations by seeking consent orders or issuing complaints, allowing for settlement of allegations.[647]

The FTC brings cases in federal court and adjudicative proceedings before administrative law judges.[648] To enforce civil penalties or seek redress, the FTC must pursue litigation in court.[649] Therefore, judicial enforcement is advantageous. When the FTC issues a complaint and an advertiser contests the allegations, the parties may proceed with an administrative trial, resulting in a judge's

[641] *See* Hirsch, *supra* note 528, at 499 ("Courts that have reviewed Section 5 and its legislative history have consistently reinforced the idea that Section 5 unfairness is broad, flexible, and capable of addressing new business practices and harms."); HOOFNAGLE, *supra* note 614, at 130 (explaining that the FTC has broad power to prevent unfair trade).

[642] Waldman, *supra* note 619, at 796.

[643] The FTC is empowered "[t]o gather and compile information concerning, and to investigate from time to time the organization, business, conduct, practices, and management of any person, partnership, or corporation engaged in or whose business affects commerce[.]" 15 U.S.C. § 46(a); *see* HOOFNAGLE, *supra* note 614, at 102–03.

[644] Van Loo proposed to expand regulatory monitoring of platforms and business information to enhance users' personal privacy and mitigate risks of data misuse. *See* Van Loo, *supra* note 603, at 1566 ("Most notably today, federal regulators do not regularly monitor the companies that run platforms, defined as sites 'where interactions are materially and algorithmically intermediated.'").

[645] To submit a consumer complaint to the FTC, see *Report Fraud to the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/faq/consumer-protection/submit-consumer-complaint-ftc [https://perma.cc/XT95-7KZP].

[646] *See* HOOFNAGLE, *supra* note 614, at 103.

[647] *See* Bladow, *supra* note 204, at 1142–43.

[648] *See* HOOFNAGLE, *supra* note 614, at 109.

[649] *See id.*; Bladow, *supra* note 204, at 1143.

recommendation to enter a cease and desist order.[650] An advertiser can be held civilly liable for up to $40,000 per violation of a cease and desist order.[651] Once the order is final, the FTC can hold a non-party liable for committing a deceptive act that violates the order, and thus can hold intermediaries responsible.[652]

A second remedial path, based on Professor Balkin's proposal, is private litigation under defamation law for an intermediary's failure to remove a false advertisement upon notification.[653] In such cases the intermediary would be held responsible under knowledge-based distributor liability for false advertisements. Whereas FTC enforcement applies to false and misleading advertisements, liability in civil litigation under defamation laws applies only when the falsehood reaches the level of defamation.[654] However, other causes of action, such as negligence, might offer remedies to individuals that prove the dissemination of falsehood caused them legally recognized harm.[655]

CONCLUSION

Online intermediaries are the governors of speech.[656] They provide communication tools and moderate the flow of information with insufficient transparency. Intermediaries are not just hosts and moderators; they recommend and target specific types of content. They profit from amplifying lies and providing targeting tools, allowing political operatives and other stakeholders to engage in a new level of information warfare.[657] Data-driven business models

[650] *See* Bladow, *supra* note 204, at 1143.

[651] *See id.*

[652] *See id.* ("Once a cease and desist order is final, the FTC can hold a nonparty liable for committing a deceptive act that violates the order.").

[653] Balkin, *supra* note 12.

[654] In the case of political advertisements, the political candidate proves the standard of actual malice and the intermediary does not benefit from defamation law defenses. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).

[655] For expansion on intermediary duty of care for design (that can be applicable regarding targeting algorithms), see Sylvain, *supra* note 308 and accompanying text.

[656] *See* Klonick, *supra* note 109, at 1670.

[657] *See* Yaël Eisenstat, *I Worked on Political Ads at Facebook. They Profit by Manipulating Us*, WASH. POST (Nov. 4, 2019), wapo.st/2qGihK6 [https://perma.cc/9MWK-EU9A].

allow advertisers to show users a different version of the truth and manipulate users with hyper-customized advertisements full of fake news stories, turning social media into a dangerous weapon.[658] In the data-driven internet era, it becomes almost impossible to separate true from false and to engage in honest discussion on matters of public importance, impinging upon the general public interest, impairing the political security of citizens, and eroding democracy.

Section 230 of the CDA immunizes intermediaries for content created by other content providers, reflecting the *internet exceptionalism* approach, which differentiates between the internet and other media that preceded it.[659] However, as technologies advance, the role of intermediaries becomes a fundamental aspect of any platform. With the transition from an internet society to an algorithmic society, intermediaries' duties should be reconsidered. This Article argues that the overall immunity regime should be nuanced, as a one-size-fits-all approach to liability is inappropriate. It endeavors to contextualize internet exceptionalism, target exceptions to overall immunity, and refine immunity to different roles intermediaries fulfill. This Article provides guidelines for deciding when immunity should apply and when it should not. It also proposes to subject intermediaries to complementary duties that promote accountability in shaping the flow of information, such as transparency obligations for moderators and algorithmic impact assessments as part of consumer protection regulation enforced by the FTC.

It is particularly important to reevaluate intermediaries' roles and duties, provide guidelines for the scope of immunity, and outline complementary transparency obligations—especially in light of recent attacks on Trump's Executive Order that strives to abolish the editorial discretion of intermediaries altogether.[660] Rethinking the scope of immunity is also important in considering recent proposals in scholarship to narrow the immunity provided for sophisticated, algorithmic-based technologies that structure, sort, target, and sell

[658] *See id.*

[659] *See* Tremble, *supra* note 575, at 847.

[660] *See* Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (June 2, 2020), *repealed by* Exec. Order No. 14,029, 86 Fed. Reg. 27,025 (May 14, 2021).

user data, thereby shaping the flow of information.[661] As Section 230's immunity gradually erodes and judicial orders continue to result in inconsistent decisions,[662] nuanced and clearer guidelines for applying immunity are necessary. This Article therefore concludes with a call for courts, policymakers, and legislators to adopt such a nuanced framework for immunity.

[661] *See* Kim, *supra* note 29, at 927; Sylvain, *supra* note 308, at 218; Tremble, *supra* note 575, at 829.

[662] *See supra* Part II.A.