TARGETING CHILDREN: LIABILITY FOR ALGORITHMIC RECOMMENDATIONS

MICHAL LAVI[*]

We live in the algorithmic society, characterized by massive digital surveillance and data collection by private companies exploiting human information vulnerabilities for profit. The infrastructure of free expression translates into an infrastructure of digital surveillance. This model, dubbed "surveillance capitalism," includes massive personalized algorithmic targeting that departs from human speakers, allowing a level of influence never witnessed before in scale, scope, or depth.

Personalized targeting can garner many benefits, as it allows individuals to find content that interests them without needing to invest energy in seeking out that content. However, personalized targeting can also cause tremendous harm because it can alter an individual's perceptual, emotional, and social judgment. It can also promote the illegal sale of firearms and drugs, increase pedophile networks, and amplify incitement to terror. Most critically for this Article, personalized algorithmic recommendations often target material to children that a platform's official policy forbids. Because children are particularly vulnerable and lack full decision-making capacity, unregulated targeting can lead to self-harm and unfortunately, has even cost children's lives. Should intermediaries

[*] Research Fellow, Hadar Jabotinsky Center for Interdisciplinary Research of Financial Markets, Crises and Technology. Research Fellow, *School of Law, Zefat Academic College*. Ph.D. (Law). The Author thanks Emily Cooper for helpful comments and excellent editorial work. Special thanks are due to Felicia D. Sych, Senior *Symposium* Editor of the *American University Law Review*, along with Paulina Andrews, Jacob Lewine, Eden Hankin, Lillian Holmes, Henessy Guerrero, Lauren Macievic, Griffen Dresner, Kristin Rheins and their colleagues on the *Law Review*'s staff for great dedication, remarkable feedback, comments, suggestions, and outstanding editorial work that profoundly improved the quality of this Article. This Article is dedicated 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.

bear any liability for targeting children? Or rather, should they be immune for autonomous targeting by artificial intelligence algorithms?

In the wake of the recent U.S. Supreme Court decisions in Twitter, Inc. v. Taamneh *and* Gonzalez v. Google*, a change in policy regarding targeting is more urgent than ever, especially in the context of targeting susceptible children. Accordingly, this Article argues that with the shift from human information to a data network connection in the algorithmic society, algorithmic targeting should be more regulated. After all, platforms are not just middlemen; algorithmic targeting differs from merely hosting content, and it is not the same as human speech. In fact, the design of the platform and the algorithmic targeting mechanism itself develop content and transform the context of the information. Therefore, intermediaries should be subject to liability for negligent design.*

This Article demonstrates how social media platforms, aiming to increase traffic and enhance profits, deliberately use algorithms to target susceptible children. This Article makes the case for imposing civil liability on intermediaries for negligent design, thereby incentivizing intermediaries to restrict algorithmic targeting directed towards children. Subsequently, this Article reviews normative free speech considerations in imposing liability for targeting children and overviews judicial decisions regarding liability by analyzing and criticizing case law interpreting § 230 immunity.

This Article outlines a proposed framework and policy reform for a duty of care for targeting children. Finally, it addresses possible objections to liability for algorithmic targeting, the problem of demonstrating causal connection between targeting and harm, and First Amendment objections to the proposed framework. The Article concludes that liability alone will not be enough to protect children, and more comprehensive reforms are required.

TABLE OF CONTENTS

INTRODUCTION

Bella Herndon was days away from turning sixteen when her parents found her hanging in her bedroom closet after she watched the first season of *13 Reasons Why*.[1]

Based on the best-selling novel, *13 Reasons Why* is a television show released on the Netflix streaming service in March 2017.[2] The show involves a high school student who dies by suicide and leaves behind thirteen cassette tapes revealing thirteen reasons why she took her own life.[3] When the show first aired, the National Association of School Psychologists issued a warning: "We do not recommend that vulnerable youth, especially those who have any degree of suicidal ideation, watch this series."[4] Moreover, suicide-prevention experts warned Netflix that the show encourages suicide, but Netflix ignored the warning and made no attempt to avoid recommending the show to youths.[5]

According to a study published in the Journal of the American Academy of Child and Adolescent Psychiatry, in the month following the show's release, there was a 28.9% increase in suicide among Americans aged ten to seventeen.[6] The National Institute of Mental

[1] Adriana Diaz, *Judge Throws out Netflix Lawsuit Brought by Father Whose 15-Year-Old Daughter Killed Herself Days After Watching Suicide Scene in* 13 Reasons Why *Because It 'Infringes on Protected Speech'*, DAILY MAIL (Jan. 12, 2022, 11:53 AM), https://www.dailymail.co.uk/news/article-10394565/Lawsuit-against-Netflix-suicide-allegedly-triggered-13-Reasons-dismissed.html [https://perma.cc/UM5Z-E259].

[2] Rachel Conte, *One Thousand Reasons: Netflix Faces Lawsuit over Failure to Warn*, BROOK. SPORTS & ENT. L. BLOG (Sept. 30, 2021), https://sports-entertainment.brooklaw.edu/film-tv/one-thousand-reasons-netflix-faces-lawsuit-over-failure-to-warn [https://perma.cc/D5U6-QTG8].

[3] Kayt Sukel, *How* 13 Reasons Why *Is Changing the Conversation About Suicide*, HARTGROVE BEHAV. HEALTH SYS. (May 25, 2017), https://hartgrovehospital.com/13-reasons-why-changing-conversation-suicide [https://perma.cc/Z5GN-3YQ6].

[4] 13 Reasons Why *Netflix Series: Considerations for Educators*, NAT'L ASS'N OF SCH. PSYCHS. (2017), https://did.li/VrECN [https://perma.cc/C569-FS9N].

[5] Benjamin Lee, *Teen Suicides Rose After Netflix's* 13 Reasons Why *Aired, US Study Shows*, GUARDIAN (Apr. 30, 2019, 2:33 PM), https://www.theguardian.com/tv-and-radio/2019/apr/30/teen-suicides-netflix-13-reasons-why-study [https://perma.cc/XH7S-TV5N].

[6] Jeffrey A. Bridge, Joel B. Greenhouse, Donna Ruch, Jack Stevens, John Ackerman & Arielle H. Sheftall et al., *Association Between the Release of Netflix's* 13 Reasons Why *and Suicide Rates in the United States: An Interrupted Time Series Analysis*, 59 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 236, 236–43 (2020).

Health attributed the spike in the adolescent suicide rate to the show.[7] Nevertheless, Netflix continued targeting the show to minors through its algorithm.[8] Following global outrage, Netflix added a "viewer warning card" before the first episode and several other graphic episodes and published a website offering information about the show and mental health resources.[9] Ultimately, in July 2019, only after public outcry did Netflix remove the graphic suicide scene from the show.[10]

Jon Herndon, Bella's father, filed a class action on behalf of Bella's estate and others similarly situated.[11] Herndon filed the action not because Netflix created and exhibited a show glorifying suicide, but because Netflix allegedly used its trove of individualized data to target children and "manipulate them into watching content that was deeply harmful to them—despite dire warnings about the likely and foreseeable consequences to such children."[12]

The U.S. District Court for the Northern District of California ignored the context created by targeting children[13] and dismissed the case, reasoning that the case arose from a protected activity—"the creation and dissemination of [the] show."[14] Since no strict liability exists for books, movies, or other forms of media, Netflix did not owe a duty to the plaintiff, an element required for negligence cases as a matter of law.[15] Herndon appealed to the Ninth Circuit, where a three-

[7] Press Release, Nat'l Inst. of Mental Health, Release of "*13 Reasons Why*" Associated with Increase in Youth Suicide Rates (Apr. 29, 2019), https://did.li/yxnrl [https://perma.cc/KP2S-NWB2].

[8] *See* Diaz, *supra* note 1 (reprimanding Netflix for continuing to target vulnerable youth).

[9] Matthew S. Schwartz, *Teen Suicide Spiked After Debut of Netflix's* '13 Reasons Why,' *Study Says*, NPR (Apr. 30, 2019, 6:20 AM), https://www.npr.org/2019/04/30/718529255/teen-suicide-spiked-after-debut-of-netflixs-13-reasons-why-report-says [https://perma.cc/B8NL-XYK5].

[10] Conte, *supra* note 2.

[11] Amended Complaint & Demand for Jury Trial, Estate of Herndon v. Netflix, Inc., No. 4:21-cv-06561, 2022 WL 551701 (N.D. Cal. Sept. 22, 2021).

[12] *Id.* ¶ 26.

[13] *See infra* Section I.D (explaining targeting of children as a "special context").

[14] Estate of B.H. v. Netflix, Inc., No. 4:21-cv-06561-YGR, 2022 WL 551701, at *3–4 (N.D. Cal. Jan. 12, 2022), *aff'd*, Estate of Herndon v. Netflix, Inc., No. 22-15260, 2024 WL 808797 (9th Cir. Feb. 27, 2024) (concluding that wrongful death actions must be brought within two years, while Herndon lodged his complaint more than four years after his daughter's suicide); Henrik Nilsson, *Netflix Keeps Win over* '13 Reasons' *Suicide Suit at 9th Circ.*, LAW360 LEGAL NEWS (Feb. 27, 2024, 8:43 PM), https://www.law360.co.uk/articles/1807395 [https://perma.cc/E8HL-UHH3].

[15] *Estate of B.H.*, 2022 WL 551701, at *3.

judge panel refused to revive the case on procedural grounds.[16] Unfortunately, this is not the first time such a tragedy occurred following the targeting of children.[17] The following stories are just a few of those examples.

Alexis Spence started using Instagram on her phone when she was eleven years old.[18] She created an online account to play with Webkinz, a stuffed animal with an online video game counterpart.[19] Soon, however, Instagram started targeting her with content that encouraged eating disorders.[20] Subsequently, her mental health deteriorated and she developed several eating disorders.[21] The Social Media Victims Law Center filed a lawsuit on behalf of Alexis and her parents against Meta Platforms, Inc. (Meta), contending that Alexis's disorders were due to "exposure to and use of Meta's unreasonably dangerous and defective social media product, Instagram."[22] The Complaint alleged that although Meta knew children under thirteen accessed its products despite its age restrictions, Meta nevertheless used targeting to expose Alexis to harmful content.[23]

Englyn Roberts died in September 2020 after she watched a video on Instagram of a woman pretending to hang herself and copied the video.[24] Englyn's parents filed lawsuits against Meta, Facebook, Inc., Snap, Inc., TikTok, Inc., and ByteDance, Inc., contending that these

[16] *Estate of Herndon*, 2024 WL 808797, at *1. Due disclosure: I submitted an Amicus Brief in support of Herndon. *See* Brief for Dr. Michal Lavi as Amicus Curiae in Support of Appellants, *Estate of Herndon*, (No. 22-15260), 2024 WL 808797 [hereinafter Dr. Lavi Brief of Amicus Curiae].

[17] *See* Clare Morell, *Social Media and Harm to Children*, ETHICS & PUB. POL'Y CTR. (Aug. 31, 2023), https://eppc.org/publication/social-media-and-harm-to-children [https://perma.cc/2UQZ-WASR] (identifying social media and targeted media as a leading driver of depression, anxiety, eating disorders, self-harm, and suicide in children).

[18] Kristin Thorne, *Long Island Family Sues Meta for 'Harming' Daughter Through Instagram Use*, EYEWITNESS NEWS (Jan. 20, 2023), https://abc7ny.com/teens-instagram-lawsuit-long-island/12718879 [https://perma.cc/G5ZS-3SSN].

[19] *Id.*

[20] *Id.*

[21] *Id.*

[22] Complaint for Personal Injuries & Jury Demand ¶ 3, Spence v. Meta Platforms, Inc., No. 3:22-cv-03294, 2022 WL 2101825 (N.D. Cal. filed June 6, 2022).

[23] *Id.* ¶¶ 3–4.

[24] Kolbe Nelson, *Teen Watched Simulated Hanging Video on Instagram Before Suicide*, CBS NEWS: 60 MINUTES OVERTIME (Dec. 11, 2022, 6:58 PM), https://www.cbsnews.com/news/instagram-hanging-video-suicide-60-minutes-2022-12-11 [https://perma.cc/63NU-3E99].

social media platforms employed algorithms targeting youth with themes of addiction and use of illegal products.[25] The plaintiffs further argued that TikTok used its products "to the detriment of [its] minor users" and steered violent videos to young subscribers.[26] For that reason, they asserted that TikTok was responsible for the death of their fourteen-year-old child.[27]

In February 2022, Chase Nasca took his own life at age sixteen after TikTok's algorithm targeted him with harmful content.[28] In an effort to understand why Chase took his life, Chase's mother turned to her son's TikTok account and found he had bookmarked, liked, saved, or favorited more than 3,000 videos.[29] Although the terms he searched were unrelated to suicide, the algorithm nevertheless recommended many videos about depression, hopelessness, and death.[30] Chase's parents sued TikTok,[31] arguing Chase did not search for the content, but rather the platform "purposefully sent [Chase] more than 1,000 videos promoting suicide, hopelessness, and self-harm in order to maximize his level of engagement on the platform," thereby boosting its advertisement revenue.[32] The plaintiffs also alleged TikTok directed

[25] Complaint for Wrongful Death and Survivorship ¶ 4, Roberts v. Meta Platforms, Inc., No. 3:22-cv-04210 (N.D. Cal. filed July 20, 2022).

[26] *Id.*

[27] Evan Peng, *TikTok Algorithm Pushes Violent Videos to Minorities, Lawsuit Says*, BNN BLOOMBERG (July 20, 2022), https://www.bnnbloomberg.ca/tiktok-algorithm-pushes-violent-videos-to-minorities-lawsuit-says-1.1794735 [https://perma.cc/U7PC-5X5D].

[28] *See* Olivia Carville, *TikTok's Algorithm Keeps Pushing Suicide to Vulnerable Kids*, BLOOMBERG (Apr. 20, 2023, 6:27 PM), https://www.bloomberg.com/news/features/2023-04-20/tiktok-effects-on-mental-health-in-focus-after-teen-suicide [https://perma.cc/TJC5-37X8] (describing Chase's Tik Tok For You feed as an "endless stream of clips about unrequited love, hopelessness, pain and what many posts glorify as the ultimate escape: suicide," including videos saying, "Take the pain away. Death is a gift," and another in which "a male voice says, 'I'm going to put a shotgun in my mouth and blow the brains out the back of my head,' and a female voice responds: 'Cool'").

[29] *Id.*

[30] *Id.*

[31] *See* Nasca v. ByteDance, Ltd., No. 23-CV-02786 (NGG), 2023 WL 5979210, at *1 (E.D.N.Y. July 27, 2023), *report and recommendation adopted*, No. 23-CV-2786 (NGG), 2023 WL 7102396 (E.D.N.Y. Oct. 27, 2023) (alleging product liability and negligence).

[32] *Social Media Victims Law Center Sues ByteDance and TikTok in the Death of 16-Year-Old Chase Nasca; Parents Travel to Washington D.C. to Hear Congressional Testimony of TikTok CEO*, BUS. WIRE (Mar. 22, 2023, 11:57 AM), https://www.businesswire.com/news/home/20230321005908/en/Social-Media-Victims-Law-Center-Sues-ByteDance-and-TikTok-in-the-Death-of-16-Year-Old-Chase-Nasca-Parents-Travel-to-Washington-D.C.-to-Hear-Congressional-Testimony-of-TikTok-CEO [https://perma.cc/354R-VMAQ].

Chase to adult accounts containing depressing and violent content;[33] moreover, TikTok even helped create these accounts by "suggesting dark, suicide-themed songs they could use to make their videos more impactful, as well as trending hashtags they could add" to maximize amplification based on TikTok's programming.[34]

Nylah Anderson died after attempting the TikTok "Blackout Challenge," which challenged users to record and post a video strangling themselves using household objects.[35] The "Blackout Challenge" is just one of many other "challenges" prevalent on TikTok that "promote dangerous behavior."[36] Nylah was hanging from a purse strap when her mother found her, and she died after several days in intensive care.[37] From January to July 2021, several other children died while attempting the same challenge.[38]

Nylah's mother sued TikTok for recommending inappropriate, dangerous, and deadly videos to users, alleging TikTok knew its algorithm encouraged children to try the challenge, yet it continued to use the algorithm anyway.[39] Although the case was ultimately dismissed,[40] TikTok continues to face pending lawsuits for the wrongful deaths of other children who participated in the "Blackout Challenge."[41] In one recent case, the California Superior Court ruled that social media companies may be held liable based on allegations of

[33] Abigail Adams, *Parents Suing TikTok over Teen Son's Death Make Emotional Appearance at Congressional Hearing*, PEOPLE (Mar. 24, 2023, 1:40 PM), https://people.com/human-interest/parents-suing-tiktok-over-death-teen-son-emotional-appearance-hearing-congress [https://perma.cc/MA6D-PU5V].

[34] *Id.*

[35] Anderson v. TikTok, Inc., 637 F. Supp. 3d 276, 278 (E.D. Pa. 2022).

[36] *Id.*

[37] *Id.*

[38] *See* Complaint at 14–15, *Anderson*, 637 F. Supp. 3d 276 (No. 2:22-cv-01849-PD) (listing four deaths associated with the "Blackout Challenge").

[39] *See id.* at 15 (alleging that TikTok "unquestionably" knew about the "Blackout Challenge" after the previous deaths).

[40] *Anderson*, 637 F. Supp. 3d at 278. This Article will expand on this case *infra* Part II.

[41] Complaint at 2, Smith v. TikTok Inc., No.22STCV21355 (Cal. Super. Ct. L.A. Cnty. filed June 30, 2022); *see* Peter Henderson, Tatsunori Hashimoto & Mark Lemley, *Where's the Liability in Harmful AI Speech?*, 3 J. FREE SPEECH L. 589, 620 (2023) (explaining how § 230 immunized TikTok from wrongful death suits originating from a self-asphyxiation challenge).

direct liability in negligence, as opposed to cases that base liability on content viewed by plaintiffs.[42]

Social media intermediaries amplify and push content to susceptible users, including content prohibited by their terms of service or community guidelines.[43] In other words, a platform can ban specific types of content while simultaneously promoting such content by algorithmic targeting,[44] which can have extremely grave consequences.[45] In 2017, for example, British teenager Molly Russell started to search images of suicide and self-harm online.[46] Pinterest, her favorite social media platform, targeted her with additional images.[47] Following such targeting, Molly died by suicide.[48] Even in the months after her death, Pinterest's algorithm continued to send her emails recommending images of graphic self-harm.[49]

In their quest to enhance profits from content and advertisement, social media intermediaries personalize content through automatic algorithms that recommend content to users.[50] This model, dubbed

[42] Joel Rosenblatt, *Kids Suing Social Media over Addiction Find a Win amid Losses*, THE STAR (Oct. 16, 2023, 9:00 AM), https://www.thestar.com.my/tech/tech-news/2023/10/16/kids-suing-social-media-over-addiction-find-a-win-amid-losses [https://perma.cc/8KL3-Y5HN].

[43] Amy B. Cyphert & Jena T. Martin, *"A Change is Gonna Come:" Developing a Liability Framework for Social Media Algorithmic Amplification*, 13 U.C. IRVINE L. REV. 155, 158 (2022).

[44] *See id.* ("Despite Prof[essor] Putnam repeatedly flagging content as objectionable, Facebook's recommendation algorithm suggests that she join a stream of similar groups—groups targeting young children and having the hallmarks of a trafficking scheme."); Michal Lavi, *Do Platforms Kill?*, 43 HARV. J.L. & PUB. POL'Y 477, 503–04 (2020) [hereinafter Lavi, *Do Platforms Kill?*] (discussing content bans in the context of conspiracy theories); Ysabel Gerrard, *Beyond the Hashtag: Circumventing Content Moderation on Social Media*, 20 NEW MEDIA & SOC'Y 4492, 4497–98 (2018) (discussing content bans in the context of eating disorders).

[45] *See, e.g.*, Ysabel Gerrard & Tarleton Gillespie, *When Algorithms Think You Want to Die*, WIRED (Feb. 21, 2019, 12:41 PM), https://www.wired.com/story/when-algorithms-think-you-want-to-die [https://perma.cc/5RNT-C93D] (explaining how eating disorders are magnified by algorithmic content despite rules against such content).

[46] *Id.*

[47] *Id.*

[48] *Id.*

[49] *Id.*

[50] Cyphert & Martin, *supra* note 43, at 165. These algorithms often result in detrimental behaviors. *See id.* ("When you're in the business of maximizing engagement, you're not interested in truth. You're not interested in harm,

"surveillance capitalism," is based on data collected from users.[51] In short, by drawing conclusions based on past interactions between the user and other users with similar profiles,[52] intermediaries exploit the unique biases of every *specific* user by using this data to target them with personalized experiences and content.[53] Personalizing content does not offer equal choice to all users, rather the algorithm determines what recommendations and content are available to whom;[54] it often targets content to susceptible users, namely children, precisely when they are most vulnerable.[55]

There are numerous examples of intermediaries using targeting to amplify harmful content to vulnerable audiences;[56] for example, about seven years ago, the New Zealand Herald revealed leaked documents demonstrating how Facebook (Meta) monitored posts, comments, and interactions on the site and how it gathered and analyzed the information to determine when minors were feeling "defeated," "overwhelmed," "stressed," "anxious," "nervous," "stupid," "silly," "useless," and a "failure."[57] By exploiting minors' moods and

divisiveness, conspiracy. In fact, those are your friends." (quoting Karen Hao, *How Facebook Got Addicted to Spreading Misinformation*, MIT TECH. REV. (Mar. 11, 2021), https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation [https://perma.cc/9MV9-YFQD])); Nancy S. Kim, *Beyond Section 230 Liability for Facebook*, 96 ST. JOHN'S L. REV. 353, 355 (2022) [hereinafter Kim, *Beyond Section 230 Liability*].

[51] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 14 (2019) (coining the term "surveillance capitalism" and explaining its impact on commerce, free will, and society).

[52] Lavi, *Do Platforms Kill?*, *supra* note 44, at 485.

[53] Michal Lavi, *Manipulating, Lying, and Engineering the Future*, 33 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 221, 231 (2023) [hereinafter Lavi, *Manipulating*].

[54] Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy & Pádraig Cunningham, *Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems*, 33 SOC. SCI. COMPUT. REV. 459, 460 (2014).

[55] Lavi, *Do Platforms Kill?*, *supra* note 44, at 553; O'Callaghan et al., *supra* note 54, at 460; *see also* Kevin Roose, *The Making of a YouTube Radical*, N.Y. TIMES (June 8, 2019), https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html [https://perma.cc/2MVX-877H] (explaining how far right YouTube creators, promoted by the algorithm, target specific demographics).

[56] *E.g.*, Nick Whigham, *Leaked Document Reveals Facebook Conducted Research to Target Emotionally Vulnerable and Insecure Youth*, N.Z. HERALD (May 1, 2017, 3:38 PM), https://www.nzherald.co.nz/business/leaked-document-reveals-facebook-conducted-research-to-target-emotionally-vulnerable-and-insecure-youth/CVTHSGXCQ4KVQCS3U6RZZ5CUJA [https://perma.cc/Z3TW-WBGA].

[57] *Id.*

insecurities, Facebook targeted advertisements to them when its algorithm believed they were most vulnerable.[58] Instagram's algorithm promoted accounts like "Prettily Skinny" and "Wanna Be Skinny" to teens looking for weight loss and dieting content.[59] Instagram has allowed underage users to receive promotional advertisements from alcohol brands.[60] YouTube has recommended videos with titles like "How to Self-Harm Tutorial" to young children.[61] Targeted algorithmic-based recommendations create a feedback loop that reinforces itself and increases the likelihood of influencing users,[62] which is all the more true when the targeted individuals are children.[63]

[58] *Id.*

[59] Donie O'Sullivan, Clare Duffy & Sarah Jorgensen, *Instagram Promoted Pages Glorifying Eating Disorders to Teen Accounts*, CNN (Oct. 4, 2021, 7:28 PM), https://www.cnn.com/2021/10/04/tech/instagram-facebook-eating-disorders/index.html [https://perma.cc/99AY-G9LH]; Dr. Lavi Brief of Amicus Curiae, *supra* note 16, at 9 n.20 ("The fictitious 13-year-old girl's Instagram account was bombarded with recommendations to follow more and more extreme dieting accounts, which could confirm and encourage self-harming inclinations and lead to eating disorders in a vulnerable young teenager." (quoting Kim, *Beyond Section 230 Liability*, *supra* note 50, at 363)).

[60] Adam E. Barry, Austin M. Bates, Olufunto Olusanya, Cystal E. Vinal, Emily Martin & Janiene E. Peoples et al., *Alcohol Marketing on Twitter and Instagram: Evidence of Directly Advertising to Youth/Adolescents*, 51 ALCOHOL & ALCOHOLISM 487, 490 (2016).

[61] Daniyal Malik, *YouTube Faces Severe Criticism for Recommending Self Harm Videos Again*, DIGIT. INFO. WORLD (Feb. 7, 2019, 2:11 PM), https://www.digitalinformationworld.com/2019/02/youtube-recommending-self-harm-videos-in-search-results-criticized.html [https://perma.cc/L5G8-9QND].

[62] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 85 (2019) (explaining the "feedback loop"); Lavi, *Do Platforms Kill?*, *supra* note 44, at 486 n.41 (citing MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 18 (2017) (describing "'persuasion sequences' of videos where insight from each one would both update someone's views and motivate them to watch another video about a related topic where they were likely to be further convinced")).

[63] Michal Gilad, Diana H. Fishbein, Gideon Nave & Nizan Geslevich Packin, *Science for Policy to Protect Children in Cyberspace*, 379 SCI. 1294, 1294–95 (2023). It should be noted that the metaverse, which is expected to be mediated through virtual spaces and augmented reality, is likely to bolster the influence of such targeting. *See* Scott Bloomberg, *Political Advertising in Virtual Reality*, 21 FIRST AMEND. L. REV. 167, 169–70 (2023) (discussing biometric monitoring in the metaverse and its possible impact on advertising); Jon M. Garon, *Legal Implications of a Ubiquitous Metaverse and a Web3 Future*, 106 MARQ. L. REV. 163, 213–14 (2022) (noting that companies may self-promote in the metaverse, raising unfair competition and consumer protection concerns); Leon Yehuda Anidjar, Nizan Geslevich Packin & Argyri Panezi, *The Matrix of Privacy: Data*

Against this background, there is an ever-increasing need to consider imposing legal liability for targeting susceptible users. This Article focuses on children because they are particularly sensitive to the influence of digital technology and are more likely to act upon the recommendations platforms targeted to them.[64] Nevertheless, intermediaries use artificial intelligence (AI) algorithms to target many other susceptible groups, often through the use of commercial targeting that pushes products and agendas on consumers, deeply influencing their decision-making on matters of life and death[65] through discrimination.[66] For example, algorithms can endanger users by steering them to particular online content by pushing notifications, allowing users to connect with drug dealers,[67] or by targeting content that incites terrorism and encourages specific audiences to perpetrate terror attacks.[68] Recommendation algorithms can even promote pedophile networks and child exploitation.[69]

Intermediaries can reduce the harm of targeting but largely refrain from doing so.[70] Intermediaries often promote specific types of content or agendas based on their strategic preferences because such

Infrastructure in the AI-Powered Metaverse, HARV. L & POL'Y REV. (forthcoming) (manuscript at 7–8), https://ssrn.com/abstract=4363208 [https://perma.cc/N549-92NX] (explaining that new technology like artificial intelligence (AI) will use data not as a commodity, but as infrastructure). *See generally* Hadar Y. Jabotinsky & Michal Lavi, *Regulating the Metaverse: Reducing Diffusion of Trader Responsibility* 58 U. MICH. J.L REFORM (forthcoming 2025) [hereinafter Jabotinsky & Lavi, *Metaverse*] (expanding on the infrastructure difference between Web 2.0 and Metaverse platforms), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4753418 [https://perma.cc/X8J7-6WEY].

[64] Gilad et al., *supra* note 63, at 1294–95.

[65] *See* Lavi, *Manipulating*, *supra* note 53, at 246–47.

[66] Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. 867, 873 (2020) [hereinafter Kim, *Manipulating Opportunity*].

[67] Dyroff v. Ultimate Software Grp., No. 17-cv-05359-LB, 2017 WL 5665670, at *1–2 (N.D. Cal. Nov. 26, 2017), *aff'd*, 934 F.3d 1093 (2019); Alex S. Rifkind, Note, Dyroff v. Ultimate Software Group, Inc.*: A Reminder of the Broad Scope of § 230 Immunity*, 51 GOLDEN GATE UNIV. L. REV. 49, 54 (2021).

[68] Lavi, *Do Platforms Kill?*, *supra* note 44, at 481–83.

[69] Nick Gallagher, *Instagram's Algorithms Help Pedophiles Find Each Other: Report*, MESSENGER (June 7, 2023), https://web.archive.org/web/20240106090116/https://themessenger.com/news/instagrams-algorithms-help-pedophiles-find-each-other-report [https://perma.cc/FTG9-AYXT].

[70] *See* Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 YALE L.J.F. 475, 510 (2021) ("[M]ost internet companies have the formidable capacity to redress these practices, but do not do so until they are called to task pursuant to a court order or an explosive news report.").

targeting triggers strong emotional responses, enhancing users' engagement with the platform and increasing profits.[71] Hence, intermediaries are reluctant to avoid such targeting.[72] According to testimony by former Facebook product manager Frances Haugen before the Senate Commerce subcommittee, Facebook knows that its algorithm promotes harmful content, yet it still resists deploying counter-measures.[73]

Although it may appear that targeting systems operate without human intervention, humans ultimately decide how,[74] when, and for what purpose to use the algorithms;[75] furthermore, humans design the algorithms and connect them to the platforms.[76] In other words, algorithm operation depends on the programmer's discretion.[77] While programmers can limit algorithmic learning processes and functions and teach the algorithm to detect, measure, and mitigate the harmful consequences of its usage, algorithms are simply there to facilitate and enable social interactions between humans.[78] Thus, even though

[71] Lavi, *Do Platforms Kill?*, *supra* note 44, at 501.

[72] Allison Zakon, *Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 2020 WIS. L. REV. 1107, 1112 (2020); Lavi, *Do Platforms Kill?*, *supra* note 44, at 500–01 ("To promote engagement, intermediaries make their website 'sticky' causing users to become addicted to the engagement and keeping them on the website."); Rupert Neate, *Extremists Made £250,000 from Ads for UK Brands on Google, Say Experts*, GUARDIAN (Mar. 17, 2017, 1:30 PM), https://www.theguardian.com/technology/2017/mar/17/extremists-ads-uk-brands-google-wagdi-ghoneim [https://perma.cc/JU8W-936C].

[73] 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/EK2L-HG5W] (last updated June 23, 2023).

[74] Derek E. Bambauer & Mihai Surdeanu, *Authorbots*, 3 J. FREE SPEECH L. 375, 380 (2023).

[75] HIDEYUKI MATSUMI & DANIEL J. SOLOVE, THE PREDICTION SOCIETY: AI AND THE PROBLEMS OF FORECASTING THE FUTURE 11 (Jan. 24, 2024), https://ssrn.com/abstract=4453869 [https://perma.cc/DC6W-PZZ3].

[76] *Id.*

[77] Michal Lavi, *Targeting Exceptions*, 32 FORDHAM INTELL. PROP., MEDIA & ENT. L.J. 65, 88 (2021) [hereinafter Lavi, *Targeting Exceptions*]; Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 615 (2019).

[78] Lavi, *Targeting Exceptions*, *supra* note 77, at 88; Philip S. Thomas, Bruno Castro da Silva, Andrew G. Barto, Stephen Giguere, Yuriy Brun & Emma Brunskill, *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). On the possibilities of

algorithms can self-learn, the programmers can limit the parameters for self-learning *ex-ante*[79] or block specific system results altogether.

Given the severity of the situation, it is understandable that state actors are turning to legislation and litigation to protect children from targeting and to address "habit-forming features that entice underage users to develop social media addictions."[80] On the litigation front, forty-one states filed a class action against Meta, arguing that the company deliberately designed addictive algorithms and concealed research that proved these designs harmed young users.[81] The lawsuit seeks to prohibit Meta from employing these addictive algorithms and from unlawfully collecting young users' personal data.[82]

This Article, however, focuses on direct civil suits, specifically users seeking to hold intermediaries liable for the harm that occurred because of their algorithmic targeting. Currently, children and their parents are filing more and more cases addressing addictive design

limiting the algorithms in the context of discriminatory biases, see ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 27 (2022).

[79] *See* Lavi, *Do Platforms Kill?*, *supra* note 44, at 537 (citing Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 367 (2016) ("Even if that initial programming permits or encourages the AI to alter its objectives based on subsequent experiences, those alterations will occur in accordance with the dictates of the initial programming.")); RONALD K. L. COLLINS & DAVID M. SKOVER, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE 27 (2018) ("[Apple's] Siri has her limitations by design. She avoids controversy; she shuns opinions; she sidesteps medical, legal, or spiritual counsel; she eschews criminal advice; and she prefers the precise and factual to the ambiguous and evaluative."); *see also* Hadar Y. Jabotinsky & Michal Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, 42 CARDOZO ARTS & ENT. L.J. (forthcoming 2024) (manuscript at 28) [hereinafter Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*].

[80] Dana DiFilippo, *N.J. Legislators Propose Punishing Social Media Companies for Kids' Online Addiction*, N.J. MONITOR (Feb. 22, 2023, 6:58 AM), https://newjerseymonitor.com/2023/02/22/n-j-legislators-propose-punishing-social-media-companies-for-kids-online-addiction [https://perma.cc/3YMT-H52Z].

[81] Complaint for Injunctive and Other Relief at 1, 38, 65, Arizona v. Meta Platforms, Inc., No. 4:23-cv-05448-YGR (N.D. Cal. filed Oct. 24, 2023), ECF No. 73-2 (contending Meta designed its algorithm with "harmful and psychologically manipulative . . . features" leading to anxiety and depression, but compelling users to stay on the application out of "fear of missing out on cultural and social trends").

[82] Mike Snider, *41 States Sue Meta Alleging that Instagram and Facebook Is Harmful, Addictive for Kids*, USA TODAY (Oct. 24, 2023, 2:10 PM), https://www.usatoday.com/story/tech/2023/10/24/meta-states-lawsuit-facebook-instagram-children/71300954007 [https://perma.cc/7MF7-5EF2].

liability.[83] These lawsuits raise several questions, but most relevantly, should the law provide an avenue of redress for families of children who have engaged in self-harm due to targeting? Should such targeting be considered a negligent design warranting liability? Should the law impose a duty to refrain from targeting children and protect them as it does in other contexts?[84] What free speech considerations and balances should be taken into consideration, and how should these considerations apply in the algorithmic society? How should the law interpret procedural barriers, particularly § 230 of the Communications Decency Act,[85] which immunizes platforms from content published by other content providers?[86] How should courts and legislators tackle these situations after the recent U.S. Supreme Court decisions in *Twitter, Inc. v. Taamneh*[87] and *Gonzalez v. Google LLC*,[88] which interpreted liability for algorithmic targeting in a related context?[89] How can a liability regime be designed to comply with First Amendment doctrine? This Article aims to provide answers to these questions, as well as develop a framework for imposing liability on intermediaries that target children.

This Article is the continuation of a line of thought that started in a previous work addressing the question: *Do Platforms Kill*?[90] While the

[83] For examples of the pending product liability and addictive design cases being brought by parents against social media companies, see Matthew B. Lawrence, *Public Health Law's Digital Frontier: Addictive Design, Section 230, and the Freedom of Speech*, 4 J. FREE SPEECH L. 299, 299 (2023) (citing Complaint at 1, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR, 2023 WL 2414002, at *1 (N.D. Cal. Mar. 8, 2023)) and Danny Tobey, Bennett Borden, Breanna Fields, Christopher Cullen, Kyle Kloeppel & Coran Darling, *Navigating the Digital Dilemma: Court Addresses Social Media Liability in Adolescent Addiction Litigation*, DLA PIPER (Jan. 11, 2024), https://www.dlapiper.com/en/insights/publications/2024/01/navigating-the-digital-dilemma-court-addresses-social-media-liability-in-adolescent-addiction [https://perma.cc/HU8T-LGBB].

[84] *See, e.g.*, Children's Online Privacy Protection Act (COPPA), 15 U.S.C. §§ 6501–06 (prohibiting website operators from collecting children's personal information). For a general proposal to impose a duty of care on intermediaries, see Danielle Keats Citron, *How to Fix Section 230*, 103 B.U. L. REV. 713, 753 (2023) [hereinafter Citron, *How to Fix Section 230*].

[85] 47 U.S.C. § 230.

[86] *See id.* § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.").

[87] 598 U.S. 471 (2023).

[88] 598 U.S. 617 (2023) (per curiam).

[89] *Id.* at 621–22; *Taamneh*, 598 U.S. at 501–04.

[90] Lavi, *Do Platforms Kill?*, *supra* note 44, at 483–84.

earlier work focused on liability for unlawful content that incites the target audience to commit imminent lawless action against third parties, this Article focuses on lawful content, as there is no legal prohibition against publishing violent or otherwise dangerous content.[91] However, this Article will argue that targeting children creates a special context which can make it unlawful[92] because its usage can serve as a breach of duty of care.[93] This Article provides answers to the earlier questions using the following structure:

Part I describes the practice of surveillance capitalism in the algorithmic society and its influences on users, particularly children. Part I then overviews studies establishing that children are highly vulnerable to the influence of dangerous, self-harming practices.[94] Subsequently, Part I explains why specifically targeting children transforms the context and thereby induces illegality. This Part demonstrates that imposing liability for such targeting is not novel and has already been recognized in scholarship and legislation, even before the digital era.[95] In fact, imposing liability for targeting is all the more justifiable in the algorithmic society.

Part II examines normative free speech considerations for corporate liability for algorithmic targeting. Subsequently, Part II also explores judicial decisions regarding liability for targeting.

Part III outlines a proposed framework for imposing liability for algorithmic targeting of children and highlights that the surveillance capitalism model allows intermediaries to determine exactly what

[91] *See* Ashcroft v. ACLU, 542 U.S. 656, 660 (2004) (holding that content-based prohibitions must be presumed invalid because they have the potential to be "a repressive force in the lives and thoughts of a free people").

[92] *See, e.g.*, California Age-Appropriate Design Code Act (CAADA), CAL. CIV. CODE § 1798.99.29(a)–(b) (West 2023) (asserting that businesses providing online services that children are likely to access should prioritize protecting children).

[93] *See* Citron, *How to Fix Section 230*, *supra* note 84, at 753 (arguing that Congress should reform § 230 to create a duty of care in matters involving "intimate privacy violations, cyber stalking, and cyber harassment").

[94] Gilad et al., *supra* note 63, at 1294–95.

[95] *See* Jane R. Bambauer, Saura Masconale & Simone M. Sepe, *Reckless Associations*, 36 HARV. J.L. & TECH. 487, 491 (2023) (arguing for a new tort remedy that uses modern forms of evidence, like network analysis, but is based in traditional theories balancing duties and liberties); *see also* Hamilton v. Accu-Tek, 62 F. Supp. 2d 802, 810 (E.D.N.Y. 1999) (noting the plaintiffs raise a novel argument claiming that the defendants negligently marketed and distributed handguns, thereby proximately causing the shootings at issue).

content is seen and by whom.[96] Subsequently, this Part overviews legislative bills recognizing the need to regulate targeting children.[97] This Part then turns to defining the targeting of children as a special context that establishes a duty of care.

In outlining a framework for imposing liability for algorithmic targeting, this Article also addresses possible objections to the framework. First, it addresses how the law can impose liability for activities performed by autonomous algorithms.[98] Second, it addresses the problem of proving causation, which can make it difficult for plaintiffs to prove targeting was the cause of the alleged harm. Finally, it addresses First Amendment objections to age verification as well as imposing liability when the First Amendment protects the targeted content.[99]

Ultimately, Part IV concludes that there is a need for more comprehensive reforms and approaches beyond liability *ex-post* to mitigate the harm caused by targeting. Critically, such approaches should focus on data collection and regulation of AI algorithms.[100]

[96] The surveillance capitalism model makes it possible for intermediaries to predict responses to targeting, exploit biases, and steer choices with personalized recommendations. Therefore, intermediaries can no longer argue they are neutral platforms. *See* MATSUMI & SOLOVE, *supra* note 75, at 37 (recognizing that people are subject to greater amounts of surveillance based on algorithmic predictions that, in turn, "distort[s] our ability to choose and create our own future").

[97] *See, e.g.*, Assemb. B. 2273, 2021–2022 Reg. Sess. (Cal. 2022) (to be codified in CAL. CIV. CODE § 1798.99.28) (showing that section 1(8) of the Bill explicitly refers to the need to disable profiling of children).

[98] Lavi, *Do Platforms Kill?*, *supra* note 44, at 537; *see* MATSUMI & SOLOVE, *supra* note 75, at 11 (explaining that it does not matter targeting will eventually be performed by AI algorithms because humans design the algorithms and can limit their operation at the design stage).

[99] *See* Brief of the Foundation for Individual Rights and Expression (FIRE); PEN American Center, Inc. ("PEN America"); the National Coalition Against Censorship (NCAC); and the Student Press Law Center (SPLC) as Amici Curiae in Support of Appellee Netflix, Inc. at 8–9, Estate of Herndon v. Netflix, Inc., No. 22-15260, 2024 WL 808797 (9th Cir. Feb. 27, 2024) [hereinafter Amici Curiae in Support of Netflix] (arguing that the First Amendment protects media depictions of self-harm).

[100] For further information on regulating algorithmic predictions under data protection and privacy laws, see MATSUMI & SOLOVE, *supra* note 75, at 52, and Jack M. Balkin, *Free Speech Versus the First Amendment*, 70 UCLA L. REV. 1206, 1269 (2023) [hereinafter Balkin, *Free Speech*] (suggesting that regulation should focus on reforming digital privacy, competition laws, and consumer protection laws instead of changing First Amendment doctrine).

I. SURVEILLANCE CAPITALISM, ALGORITHMIC TARGETING, SUSCEPTIBLE CHILDREN, HARM, AND LIABILITY

This Section explores the intricacies of surveillance capitalism and algorithmic targeting. In particular, this Section explains how technology companies exploit user data in an effort to boost users' engagement with the platforms.

A. Surveillance Capitalism and Targeting

> We know what you played, searched for, or rated, as well as the time, date, and device. We even track user interactions such as browsing or scrolling behavior. All that data is fed into several algorithms, each optimized for a different purpose. In a broad sense, most of our algorithms are based on the assumption that similar viewing patterns represent similar user tastes. We can use the behavior of similar users to infer your preferences.[101]

In recent decades, technology companies have drastically increased their influence through the development of the surveillance capitalism model, marking the new economic order of the twenty-first century.[102] This model relies on collecting massive amounts of user information and commercializing online experiences for economic benefit.[103] Constant private surveillance and documentation of the public's behavior is the "new oil" of commercial profit.[104] This data is constantly created by users and collected and analyzed by tech companies, particularly social media platforms.[105] The main goal of this collection is to extract users' behavioral data to better target content and advertisements, thereby increasing corporate profits.[106] These practices are the flesh and bones of the algorithmic age and depend on "pervasive surveillance and data collection."[107]

The more users consume, share, and engage on social media platforms, the more data collected and processed,[108] the more accurate

[101] *See* Tom Vanderbilt, *The Science Behind the Netflix Algorithms that Decide What You'll Watch Next*, WIRED (Aug. 7, 2013, 6:30 AM), https://www.wired.com/2013/08/qq-netflix-algorithm [https://perma.cc/F6XW-ZBZZ].

[102] ZUBOFF, *supra* note 51, at 100.

[103] Lavi, *Manipulating*, *supra* note 53, at 227.

[104] *Id.*

[105] *Id.* (noting that some technology companies offer their services in exchange for collecting and analyzing the data of their end users).

[106] *Id.* at 229.

[107] Balkin, *Free Speech*, *supra* note 100, at 1243.

[108] Lavi, *Manipulating*, *supra* note 53, at 248.

the targeting, and ultimately, the more companies' profits increase.[109] Therefore, companies strive to enhance engagement and make users stay on the platform longer.[110] To boost engagement, technology companies make their platforms "sticky" and addictive,[111] often by amplifying emotionally-charged content, like violence, through algorithmic recommendations that disregard users' well-being.[112]

Recommendation systems are employed using pervasive data collection under the surveillance capitalism model.[113] As I have demonstrated in previous work, recommendation systems are the end result of the data lifecycle,[114] but this cycle begins with vast data collection on digital product end users.[115] Social media platforms are designed to enhance data collection by seducing users into sharing more information and addicting users to their service.[116] Not only can tech companies collect data through users' active online engagement, but they can also collect data from individuals based on their everyday interactions with others who are connected to devices; such information is created automatically.[117] In the age of the Internet of Things, which merges online and offline activities, and in the wake of

[109] *Id.*

[110] Michal Lavi, *Publish, Share, Re-Tweet, and Repeat*, 54 U. MICH. J.L. REFORM 441, 460 (2021) [hereinafter Lavi, *Re-Tweet*].

[111] *See* Lawrence, *supra* note 83, at 5 ("[P]latforms have designed their apps (either knowingly or negligently) to foster compulsive use in unwitting users, including both kids and adults, with widespread and harmful effects on the public health.").

[112] Lavi, *Do Platforms Kill?*, *supra* note 44, at 501; Zakon, *supra* note 72, at 1108–09.

[113] Balkin, *Free Speech*, *supra* note 100, at 1243.

[114] *See* Lavi, *Manipulating*, *supra* note 53, at 249 (arguing that the final stage—knowledge of individual's behavior—allows companies to influence decision-making and pushes individuals to think differently than they would otherwise).

[115] *Id.* at 238.

[116] ZUBOFF, *supra* note 51, at 614 (explaining that "just as ordinary consumers can become compulsive gamblers at the hands of the gaming industry" behavioral technology draws "ordinary young people . . . into an unprecedented vortex of social information"); Catherine Price, *Trapped—The Secret Ways Social Media Is Built to Be Addictive (And What You Can Do to Fight Back)*, BBC SCI. FOCUS MAG. (Oct. 29, 2018, 4:00 AM), https://www.sciencefocus.com/future-technology/trapped-the-secret-ways-social-media-is-built-to-be-addictive-and-what-you-can-do-to-fight-back [https://perma.cc/V3NU-MAHZ] (likening social media mechanisms designed to keep users' attention to those employed by casinos); Hilary Andersson, *Social Media Apps Are 'Deliberately' Addictive to Users*, BBC (July 3, 2018), https://www.bbc.com/news/technology-44640959 [https://perma.cc/HS3B-FPKT]; Zakon, *supra* note 72, at 1114.

[117] Lavi, *Manipulating*, *supra* note 53, at 239.

the metaverse, which merges virtual reality with reality itself,[118] today, data collection operates on a larger scale than ever before and collects new types of information, such as movements, facial expressions, vocal inflections,[119] and vital signs, making it possible to predict users' emotional states[120] without their meaningful consent.[121]

The second stage of the data lifecycle is analysis and profiling.[122] Profiling involves making inferences about individuals based on the data collected on them;[123] profiles are constructed by comparing facts about a person to facts about others.[124] Modern algorithms are a game changer in the field of profiling;[125] complex algorithms mine information, find connections and correlations, draw conclusions, and can even predict users' future behavior, feelings, and thoughts.[126] Because predictions are based on large amounts of data, high collection rates, and a variety of data types, algorithmic prediction capabilities are gaining a distinct advantage over human prediction.[127]

The final stage of the information lifecycle is influencing decision-making in various ways, particularly by using personalized recommendation systems and targeting.[128] Collection and analysis of information makes it possible to target messages to susceptible audiences, thereby shaping their behavior.[129] The algorithmic society allows extraordinary and unprecedented targeting and tailoring

[118] *See* Anidjar et al., *supra* note 63, at 22; Jabotinsky & Lavi, *Metaverse*, *supra* note 63, at 34.

[119] *E.g.*, JOSEPH TUROW, THE VOICE CATCHERS: HOW MARKETERS LISTEN IN TO EXPLOIT YOUR FEELINGS, YOUR POLICY, AND YOUR WALLET 1 (2021).

[120] Lavi, *Manipulating*, *supra* note 53, at 232.

[121] For further information on the fiction of consent in the digital age, see Daniel J. Solove, *Murky Consent: An Approach to the Fictions of Consent in Privacy Law*, 104 B.U. L. REV. 593 (2024) and Daniel J. Solove, *Artificial Intelligence and Privacy*, 77 FLA L. REV. (forthcoming 2025) (manuscript at 29) [hereinafter Solove, *Artificial Intelligence and Privacy*], https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4713111 [https://perma.cc/F982-3KBP].

[122] Lavi, *Manipulating*, *supra* note 53, at 246.

[123] MATSUMI & SOLOVE, *supra* note 75, at 9.

[124] *Id.* at 9–10.

[125] *Id.* at 10.

[126] Lavi, *Manipulating*, *supra* note 53, at 246.

[127] Balkin, *Free Speech*, *supra* note 100, at 1246–49.

[128] Lavi, *Manipulating*, *supra* note 53, at 29, 249–50.

[129] Alexander Tsesis, *Marketplace of Ideas, Privacy, and the Digital Audience*, 94 NOTRE DAME L. REV. 1585, 1589–90 (2019).

capabilities.[130] Models of targeting are not just based on exploiting general insights, heuristics, and biases;[131] indeed, the new data-driven models extend beyond exploitation of collective cognitive limitations of individuals.[132] Targeting is built on personalization procured from ever-richer sources of data, thereby allowing exploitation of the unique traits of every specific individual.[133] Beyond targeting based on a general age group or location, targeting can also be based on a person's lifestyle patterns or personality traits according to psychographic profiling deduced from collected data.[134]

Targeting can also be based on a person's current mood and emotional state; companies take users' data to determine exactly when a user needs a "confidence boost."[135] Companies also detect users' emotional states by conducting various experiments involving linguistic analysis thereby allowing them to more accurately target and influence the audience to increase engagement with the platform.[136] Emotion recognition technology is developing rapidly. For instance, the video and image understanding platform Lumos can comb through and analyze photos and videos uploaded to social media platforms, learn what they contain, and then conduct sophisticated facial recognition to uniquely identify people and emotions in their facial expressions.[137] And at its simplest level, targeting can be based

[130] *See* Lavi, *Manipulating*, *supra* note 53, at 249–50 (describing methods and targets of data collection by media companies).

[131] *Id.*

[132] *Id.*

[133] *Id.* Notably, such personalized targeting surpasses exploiting existing biases because it can form new biases. *See id.* at 250–51.

[134] The five characteristics are known as the OCEAN model (an acronym of the personality traits). Hannes Grassegger & Mikael Krogerus, *The Data that Turned the World Upside Down*, VICE: MOTHERBOARD (Jan. 28, 2017, 9:15 AM), https://www.vice.com/en/article/mg9vvn/how-our-likes-helped-trump-win [https://perma.cc/28ED-HFUC] (defining Openness: the need for new experiences; Conscientiousness: whether a person prefers the status quo or needs changes; Extroversion: whether a person is friendly; Agreeableness: whether a person takes care of others and puts their needs first; and Neuroticism: whether a person tends to worry).

[135] Lavi, *Manipulating*, *supra* note 53, at 255–58; *see supra* notes 75–76 (discussing how Facebook monitored and analyzed social media posts to perfect target timing).

[136] Julie E. Cohen, *The Emergent Limbic Media System*, *in* LIFE AND THE LAW IN THE ERA OF DATA-DRIVEN AGENCY 60, 61 (Mireille Hildebrandt & Kieron O'Hara eds., 2020).

[137] Scott Berinato, *Inside Facebook's AI Workshop*, HARV. BUS. REV. (July 19, 2017), https://hbr.org/2017/07/inside-facebooks-ai-workshop [https://perma.cc/8FNB-8WQ3].

on engagement on social networks, such as users' clicks, likes, and shares, in addition to topics discussed among users.[138]

The combination of the surveillance society model and AI algorithms allows companies to conduct sophisticated targeting by creating a context of vulnerability. Specifically, by using opaque algorithms, companies use targeting to influence the intuitive, emotional, and instinctive thought process ("system 1"), while circumventing the deliberative mode of thought ("system 2").[139] For example, when an algorithm learns that a user is afraid of something, it targets this emotion by specifically recommending content that fuels this fear, thereby manipulating the user to consume subjectively harmful content.[140] Companies can also create a "framing"[141] effect by influencing social dynamics on social networks and reinforcing social pressure by emphasizing specific information posted on users' newsfeeds.[142] Companies not only react to users' emotions, they can also influence emotions to improve responsiveness to targeting.[143] The sophistication of targeting is developing rapidly, and opportunities to influence are expected to continue expanding in scale and scope.[144]

[138] Lavi, *Manipulating*, *supra* note 53, at 258–59.

[139] *Id.* at 260.

[140] Edward Muldrew, *Understanding the "YouTube Rabbit Hole"*, MEDIUM (July 26, 2019), https://medium.com/swlh/understanding-the-youtube-rabbit-hole-4d98e921eabe [https://perma.cc/NZ45-E3YR] (describing YouTube's recommendation system and the "rabbit hole" phenomenon).

[141] RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 36–37 (2008) (explaining that "framing" is the strategic presentation of information to influence decision-making which exploits the tendency of individuals to be passive decision-makers).

[142] *See* Lavi, *Manipulating*, *supra* note 53, at 261.

[143] *Id.* at 261–62. *See generally* MOSHE GLICKMAN & TALI SHAROT, HOW HUMAN-AI FEEDBACK LOOPS ALTER HUMAN PERCEPTUAL, EMOTIONAL AND SOCIAL JUDGMENTS, https://doi.org/10.31219/osf.io/c4e7r [https://perma.cc/S59U-LBM7] (demonstrating that biased AI systems can change people's perceptual, emotional, and social judgments and distort them more than ever before). Facebook previously experimented with subliminal exposure to specific emotional content by showing some users only negative posts on their newsfeed while showing another group of users only positive posts. *See* James Grimmelmann, *The Law and Ethics of Experiments on Social Media Users*, 13 COLO. TECH. L.J. 219, 221–23 (2015). This subliminal exposure led individuals to alter the tone of their posts to reflect the newsfeeds they received. Lavi, *Manipulating*, *supra* note 53, at 262. The strategies described here are just some of the strategies of influence. For more strategies of influence, see *id.* at 259–62.

[144] This is true especially in the age of the Metaverse. *See* Lavi, *Manipulating*, *supra* note 53, at 233 ("The metaverse and augmented reality present new opportunities to

Critically, algorithms take control away from users, who are unaware that they are in a bubble because algorithms are opaque and operate in a black box.[145]

B. Targeting and Harm

> The use of sophisticated technology and the engineering of its products creates a harm that is separate and distinct from any harm that the content itself might inflict . . . Facebook's algorithms feed the user's interest, and its design features encourage obsessions, physical inactivity, and other unhealthy and harmful behavior.[146]

Targeting detrimental-yet-lawful content posted by third parties poses inherent risks, especially when the target audience includes children. The practice of personalized AI algorithmic recommendations encourages the target audience to consume harmful content by controlling what content they see online based on their past activity.[147] This can be accomplished by using a model dubbed the "rabbit hole," a phenomenon of entering a website intending to watch a video or two but ultimately following algorithmic recommendations leading to endless videos,[148] each one more extreme than the previous.[149] In the age of surveillance capitalism, intermediaries operating online platforms "know" what users watch[150]

monitor users, including psychological responses and biometric data such as facial expressions.").

[145] *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015) (describing the incongruity of large corporations with opaque operating procedures amassing greater access to citizens' personal information as analogous to a "black box"); *How Filter Bubbles Isolate You*, GCF GLOB., https://edu.gcfglobal.org/en/digital-media-literacy/how-filter-bubbles-isolate-you/1 [https://perma.cc/T7DR-3SYK] ("[T]hese algorithms don't ask for your permission, tell you when they're active, or say what they're keeping from you.").

[146] Kim, *Beyond Section 230 Liability*, *supra* note 50, at 377–78.

[147] *See* O'Callaghan et al., *supra* note 54, at 460 (investigating the effects of YouTube's recommender system on vulnerable individuals); Roose, *supra* note 55 (chronicling the impact of YouTube and its recommender algorithm on one person's ideological transformation).

[148] O'Callaghan et al., *supra* note 54; Muldrew, *supra* note 140.

[149] Muldrew, *supra* note 140; *see also* O'Callaghan et al., *supra* note 54 (asserting that users who view an "extreme right" YouTube video often receive recommendations for more extreme right content, swiftly drawing them into an "ideological bubble after just a few clicks").

[150] *See generally* Nico Grant, Natasha Singer & Aaron Krolik, *YouTube Ads May Have Led to Online Tracking of Children, Research Says*, N.Y. TIMES (Aug. 17, 2023), https://ww

and they can predict users' future viewing preferences, thereby encouraging them to spend as much time as possible on the platform.[151]

This "rabbit hole" strategy reinforces user beliefs and dispositions, rendering them more extreme.[152] For example, a user might seek information or videos focusing on an effective diet, and the algorithm might lead them to videos encouraging eating disorders.[153] Ysabel Gerrard, a scholar from the *University of Sheffield*, demonstrated that after using search terms that could relate to dieting, she started receiving automatic recommendations for pro-eating disorder videos.[154] Often, targeting such content contradicts the platform's own policy prohibiting the publication of content that encourages self-harm or violence,[155] but nevertheless, the intermediary targets such content to susceptible users.[156] As demonstrated in this Article's introduction, harmful content is often targeted to children.[157] Algorithmic recommendations can create a feedback loop that reinforces itself and inflicts real harm by pushing children to self-harm.[158] Such consequences are distinct from the harm inflicted by the

w.nytimes.com/2023/08/17/technology/youtube-google-children-privacy.html [https://perma.cc/A9DN-SVB8] (reporting on a study of web-browsing-data collection demonstrating likely tracking of children's browser activity).

[151] Muldrew, *supra* note 140; ZUBOFF, *supra* note 51, at 466.

[152] Roose, *supra* note 55.

[153] Gerrard, *supra* note 44, at 4505.

[154] Lavi, *Do Platforms Kill?*, *supra* note 44, at 503 (citing Gerrard, *supra* note 44, at 4505).

[155] *Id.* at 503 (discussing Gerrard, *supra* note 44, at 4505); *see Suicide, Self-harm, and Eating Disorders* Policy, GOOGLE: YOUTUBE HELP, https://support.google.com/youtube/answer/2802245?hl=en [https://perma.cc/XW5A-9GUE].

[156] Gerrard, *supra* note 44, at 4505.

[157] *See, e.g.*, Whigham, *supra* note 56 (alleging that Facebook intentionally targeted potentially vulnerable children with predatory advertising tactics); O'Sullivan et al., *supra* note 59 (stating that Instagram and its parent company Facebook promoted accounts which "glorified" eating disorders to accounts owned by teenagers); Malik, *supra* note 61 (providing that YouTube had been recommending content containing images of self-harm to its users, some as young as thirteen years old).

[158] *See* Appellants' Opening Brief at 16–18, Estate of Herndon v. Netflix, Inc., No. 22-15260, WL 808797 (9th Cir. Feb. 27, 2023). *See generally* Complaint for Personal Injuries, Spence v. Meta Platforms, Inc., No. 3:22-cv-03294, (N.D. Cal. filed June 6, 2022) (alleging that Meta's Instagram platform was unreasonably dangerous and, therefore, liable for injuries suffered by an eleven year old Instagram user, including addiction, anxiety, depression, self-harm, eating disorders, and suicidal ideation); Complaint for Wrongful Death and Survivorship, Roberts v. Meta Platforms, Inc., No. 3:22-cv-04210, (N.D. Cal. filed July 20, 2022) (alleging strict product liability (design

content itself because this loop directs users to specific personalized content instead of a default in which users only view content they actively seek; this loop thereby creates a context of vulnerability.

In previous work, I demonstrated how the surveillance capitalism model in the algorithmic society influences consumer behavior and decision-making, and I explained how such influence reaches the level of manipulation;[159] this manipulation constitutes an *intentional* attempt to influence a subject's behavior by *exploiting a bias or vulnerability*.[160] I proposed soft regulation focusing on contextual disclosure, and I advocated for a new remedy of compensation for autonomy infringement by powerful speakers.[161] Disclosure, however, is insufficient when the targeting is directed at children, who are more susceptible to targeting and are too young to understand a statement of disclosure.[162] Because children's well-being is put at risk when platforms target them with harmful content,[163] stricter regulation is needed.

C. Targeting Children: Their Unique Susceptibility

Children are more susceptible to influence than adults[164] because their brains are not fully developed.[165] Consequently, they (1) lack the meta-awareness and media literacy skills needed to critically evaluate

defect and failure to warn) and negligence related to the effects of social media algorithms on the mental health of teenagers).

[159] *See generally* Lavi, *Manipulating*, *supra* note 53 (asserting that surveillance capitalism, driven by technology and marketing, manipulates consumer behavior, posing threats to individual autonomy, free speech, and democratic principles).

[160] Shaun B. Spencer, *The Problem of Online Manipulation*, 2020 U. ILL. L. REV. 959, 990 (2020) (emphasis added); *see also* Lavi, *Manipulating*, *supra* note 53, at 269–70 (discussing the elements of manipulation).

[161] Lavi, *Manipulating*, *supra* note 53, at 298–312 (arguing for Federal Trade Commission (FTC) enforcement of specific disclosure obligations for media companies and a private enforcement remedy for compensation for infringement of autonomy).

[162] Thomas Christiano, *Algorithms, Manipulation, and Democracy*, 52 CANADIAN J. PHIL. 109, 115 (2022).

[163] *See, e.g.*, *supra* note 158 (listing cases alleging self-harm resulting from social media use).

[164] *See* Christiano, *supra* note 162, at 115 ("[I]t is thought that children are more susceptible to manipulative advertising than adults. Adults, in significant part, seem to realize that advertising is not to be taken at face value. So, they are not manipulated by it, at least not to the extent that children are.").

[165] DANAH BOYD, IT'S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS 183–86 (2014).

the content they consume;[166] (2) are more vulnerable to peer influences;[167] and (3) are more impulsive and their decision-making abilities have not yet matured, so they are not well-equipped to make their own decisions.[168]

A broad range of scientific studies have demonstrated the substantial gap between children and adults in their biological, cognitive, and emotional development.[169] Children today are digital natives,[170] and technology is an integral part of their lives.[171] Social media platforms exploit and manipulate children's vulnerabilities and addict them to digital platforms through various features.[172] Because children have a limited capacity to discern advertising biases, advertisers and

[166] *See* Holli Sargeant, Technologies of Deception Conference at Yale Law School: A Rights-Based Approach to Online Economic Exploitation of Children 17 (Mar. 25, 2022), https://ssrn.com/abstract=4106649 [https://perma.cc/7B49-YQY9]; Marco Scalvini, *Making Sense of Responsibility: A Semio-Ethic Perspective on TikTok's Algorithmic Pluralism*, SOC. MEDIA & SOC'Y, Apr.–June 2023, at 1, 2 ("Young users may be vulnerable to the influence of algorithmic recommendation systems due to a lack of media literacy skills.").

[167] Gilad et al., *supra* note 63, at 1295.

[168] *Teen Brain: Behavior, Problem Solving, and Decision Making*, AACAP (Sept. 2017), https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/The-Teen-Brain-Behavior-Problem-Solving-and-Decision-Making-095.aspx [https://perma.cc/TN5R-XKVT] (noting that adolescents are "more likely to: []act on impulse[,]misread or misinterpret social cues and emotions . . . [and] engage in dangerous or risky behavior").

[169] Gilad et al., *supra* note 63, at 1294.

[170] JOHN PALFREY & URS GASSER, BORN DIGITAL: UNDERSTANDING THE FIRST GENERATION OF DIGITAL NATIVES 230, 292 (2008); BOYD, *supra* note 165, at 196.

[171] Gilad et al., *supra* note 63, at 1295 ("[E]mpirical surveys conducted before the pandemic showed that more than a third of children in the US began using devices while still in diapers, with the national average of daily screen-time consumption for all children aged 8 and above being between [four] and [nine] hours.").

[172] For example, social media platforms are incorporating game-like features in their platforms and services. *Id.* at 1296 (explaining that gamification is achieved by "including scoring, rewards, novelty, peer competition, game rules, quests, and challenges"). These features can lead children to make poor decisions. *See* Lydia Patrick, *TikTok 'Put Children in Danger' by Failing to Take down 'Challenge' Videos, Claim Grieving Parents*, INDEP. (Mar. 2024, 4:46 PM), https://www.independent.co.uk/news/uk/home-news/tiktok-instagram-youtube-online-challenges-b2493519.html [https://perma.cc/9KVW-MRFD] (reporting multiple instances of severe adolescent accidental self-harm resulting from TikTok challenges). Another example is embedding social features in the platforms which enable direct interaction with other users, allowing cyberbullies, or even just advertisers, to get in touch with children and negatively influence them. Gilad et al., *supra* note 63, at 1296.

commercial bodies can exploit them more easily than adults, thereby influencing their decision-making and shaping their behavior.[173]

In fact, studies have found that exposure to such manipulation can induce chemical imbalances in children's brains and alter their brains' structural and functional development.[174] These changes can interfere with children's ability to develop,[175] hinder their social skills, and induce stress, posing several health risks like increased heart rate and blood pressure.[176] Even more relevantly, platforms that target children with harmful content can, and do, result in self-harm.[177]

D. Targeting Children as a Special Context: Should the Law Impose Liability or Regulate Targeting Even if the Content Itself Is Not Unlawful?

This Section proposes that a context of vulnerability is created when intermediaries influence children through algorithmic targeting. Given the implications and risks of targeting children, this Section underscores the need for the law to impose a stricter framework for liability.

1. Targeting children as a special context

Online speech does not occur in a void; it exists in many contexts, and each context provides distinctive types of interactions among users.[178] As I wrote in a previous article on a related issue, "[t]he source of the message, the context of the message, and the situation influence the flow of information."[179] Studies reveal the context may even be more important than the content itself;[180] indeed, "[s]imple changes

[173] Gilad et al., *supra* note 63, at 1296.

[174] *Id.*

[175] *Id.*

[176] *Id.*

[177] *See supra* notes 171–76 and accompanying text (showing children's susceptibilities to manipulative designs on digital platforms, exposure to external influencers, inability to discern advertising biases, and negative impacts on brain development, social skills, and physiological health from excessive social media exposure).

[178] Michal Lavi, *Taking out of Context*, 31 HARV. J.L. & TECH. 145, 193 (2017) [hereinafter Lavi, *Taking out of Context*].

[179] *Id.*

[180] *Cf.* MALCOLM GLADWELL, THE TIPPING POINT: HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE 158 (2002) (describing a study suggesting situations cause a person's action, at least in part, rather than just inherent traits); CHARLES KADUSHIN, UNDERSTANDING SOCIAL NETWORKS: THEORIES, CONCEPTS, AND FINDINGS 146–48 (2012)

in the source of the message, the manner of presentation, and the nature of the recipients, [can] influence the magnitude and credibility ascribed to the content."[181] Intermediaries are more than mere middlemen; their algorithms pique user interest and influence what users view, value, believe, and repost.[182] They target personalized recommendations and repeatedly expose users to specific types of content, thereby altering the context of the flow of information, affecting users' moods and shaping their behavior.[183] Targeting reinforces messages and sources through repetition,[184] and it can also change social dynamics,[185] thereby creating a special context of vulnerability that can dramatically increase the content's impact. When it comes to children, targeting's influence can have far-reaching effects and can even cost lives.

2. *Targeting children and liability*

> [T]he more that digital companies boast of their abilities to predict, nudge, and influence their end users successfully, the more they appear responsible for their content moderation decisions and their recommendation algorithms.[186]

When intermediaries target unlawful content, such as incitement to commit imminent lawless action, imposing liability intuitively appeals

(discussing studies that suggest social circles exert strong influence over adolescents' opinions and decisions); HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 50 (2010) (describing how people generally seek aggregated personal information systems in certain contexts but not others); James Grimmelmann, *Privacy as Product Safety*, 19 WIDENER L.J. 793, 810–12 (2010) (discussing how personal data collection is essential to online functions, particularly Facebook, but can be dangerous in other contexts).

[181] Lavi, *Taking out of Context*, *supra* note 178, at 193.

[182] Lavi, *Targeting Exceptions*, *supra* note 77, at 73, 138; Lavi, *Re-Tweet*, *supra* note 110, at 451 ("[T]he more individuals are exposed to a particular statement, the more likely they are to believe it and perceive it as a known fact.").

[183] *See* Whigham, *supra* note 56.

[184] Lavi, *Targeting Exceptions*, *supra* note 77, at 146 ("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.' In such cases the intermediary can be perceived as the source of the message and not just a mere platform.").

[185] Lavi, *Re-Tweet*, *supra* note 110, at 499.

[186] Balkin, *Free Speech*, *supra* note 100, at 1253.

to the legal mind.[187] Although liability for targeting children is less intuitive when the content is not unlawful per se, strong justifications exist for imposing liability for such targeting as well. Given the unique concern for children's safety, some scholars propose that targeting children alters the content's legal context.[188] This change could render lawful content unlawful or generate a context of vulnerability justifying increased protection and a duty of care.[189]

Imposing liability in such circumstances is not revolutionary.[190] Indeed, there are several instances where regulation or liability of content or activities, while not unlawful per se, may be justified due to their context, presentation, or potential audience. In some circumstances, targeting or simply publishing specific content is lawful, except when the target audience is children. For example, targeting alcohol advertisements to minors creates a context of illegality under the law itself, which in turn recognizes the need to restrict such targeting.[191] Thus, many state laws specifically forbid directing alcohol advertisements toward minors.[192] Below are other examples of regulation or liability of content or activities that are not unlawful per se, yet their context or presentation and potential audience justified liability or other regulation.

In *FCC v. Pacifica Foundation*,[193] for example, the U.S. Supreme Court held that the Federal Communications Commission (FCC) could penalize a radio station that broadcasted a comedy that included profanity, although such comedy is legal in other venues, such as comedy clubs.[194] In doing so, the Court justified restrictive speech rules

[187] *See* Lavi, *Do Platforms Kill?*, *supra* note 44, at 547–67 (discussing liability already imposed upon intermediaries for targeting unlawful conduct and suggesting further legal frameworks that can provide a more nuanced liability scheme).

[188] *See id.* at 499–50, 454 (describing legal frameworks for imposing liabilities on intermediaries).

[189] *See id.*

[190] *See, e.g.*, FCC v. Pacifica Found., 438 U.S. 726, 732, 750–51 (1978) (upholding a law penalizing a radio station for broadcasting profanity).

[191] *See, e.g.*, ALA. ADMIN. CODE r. 20-X-7-.01(h) (2024) ("No advertisement shall include anything which might appeal to minors by implying that the consumption of alcoholic beverages is fashionable or the accepted course of behavior.").

[192] *Id.* For further information on state laws that restrict targeting of children, see THE CTR. ON ALCOHOL MKTG. & YOUTH, STATE ALCOHOL ADVERTISING LAWS: CURRENT STATUS AND MODEL POLICIES 4 (2003), https://www.pewtrusts.org/~/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/alcohol_marketingand_youth/hhscamystatereportpdf [https://perma.cc/PHP4-AAPL].

[193] 438 U.S. 726 (1978).

[194] *Id.* at 750–51.

for broadcasts "given radio's ability to intrude unexpectedly into the home of an unwilling listener, and given the potential presence of *children in the audience*."[195]

Indeed, in *Reno v. ACLU*,[196] a decision from 1997 concerning internet speech restrictions, the U.S. Supreme Court rejected analogies to precedents upholding restrictions on broadcasting content that could influence vulnerable audiences.[197] However, the reason for rejecting these analogies was the uniqueness of broadcast media, including the scarcity of available frequencies, which, according to the Court, were "not present in cyberspace."[198] However, in today's algorithmic society, the internet has changed and is significantly different from when the court issued its 1997 decision.[199]

In another example, a U.S. district court even went so far as to recognize liability for creating a context of vulnerability. In *Hamilton v. Accu-Tek*,[200] a New York court found three handgun manufacturers collectively liable in the shooting of Stephen Fox, a sixteen-year-old boy,[201] because they did not exercise reasonable care when marketing and distributing their handguns.[202] The court "impose[d] an affirmative duty upon handgun manufacturers to market and distribute handguns in a manner that prevents future criminal misuse."[203] Thus, even though selling guns *can* be legal, the context mattered.[204] In short, failing to prevent guns from reaching the black market is a context the law subjects to liability and considers "negligent entrustment of a dangerous instrumentality to an incompetent

[195] *See* Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard*, 1 J. FREE SPEECH L. 227, 253 (2021) (emphasis added).

[196] 521 U.S. 844 (1997).

[197] *Id.* at 868; Keller, *supra* note 195, at 253.

[198] Keller, *supra* note 195, at 253 (quoting *Reno*, 521 U.S. at 868).

[199] *Id.*

[200] 62 F. Supp. 2d 802 (E.D.N.Y. 1999), *vacated*, Hamilton v. Beretta U.S.A. Corp., 264 F.3d 21 (2d Cir. 2001).

[201] *Id.* at 811, 835.

[202] Colin K. Kelly, Hamilton v. Accu-Tek*: Collective Liability for Handgun Manufacturers in the Criminal Misuse of Handguns*, 103 W. VA. L. REV. 81, 82 (2000) (noting that because the plaintiff "could not link his injuries to any specific .25 caliber handgun manufacturer, [the court] allowed the jury to apportion liability according to each manufacturers' [sic] share of the national .25 caliber handgun market").

[203] *Id.*

[204] *See id.* at 93 (noting that complying with relevant laws and regulations is not always enough to insulate gun manufacturers from liability).

user."[205] Similarly, in addition to manufacturers facing liability for negligent practices enabling criminal gun misuse, parents of minors who committed deadly shootings at school have been held liable for not adequately securing their guns.[206]

Another case imposing liability for creating a context that facilitated harm is *Weirum v. RKO General, Inc.*[207] In this controversial ruling, the California court held liable a radio station that promised a prize to the first listener who could locate a popular radio D.J. in the street.[208] Although there was no explicit or implicit encouragement to drive unsafely to locate the D.J., the court concluded that it was foreseeable that the contest would put third parties at heightened risk.[209] Indeed, some argue the court went too far and other courts have not followed this ruling.[210] Nevertheless, if a court can impose liability for creating a context that encourages unsafety, the court is all the more justified in imposing liability when a defendant obviously exerted harmful influence.

Additional examples outside of judicial decisions illustrate the law's recognition that targeting can be problematic. Legislators have imposed liability for merely facilitating activity or selling goods that resulted in harm.[211] Laws have treated facilitating harm as negligence, even if the entity sued did not directly cause the harm.[212] For example, some laws allow a plaintiff injured by an intoxicated driver to sue both the driver *and* the party responsible for overserving the driver, even though selling alcohol to adults is legal in general.[213]

Recent literature has also recognized influence as a behavior that could be subject to liability. Legal scholars Jane R. Bambauer, Saura

[205] Bambauer et al., *supra* note 95, at 530 (quoting John C. P. Goldberg & Benjamin C. Zipursky, *The* Restatement (Third) *and the Place of Duty in Negligence Law*, 54 VAND. L. REV. 657, 683 (2001)).

[206] *Id.* at 531 (citing Jack Healy, *Behind the Charges Faced by the Parents of the Michigan Shooting Suspect*, N.Y. TIMES (Dec. 3, 2021), https://www.nytimes.com/2021/12/03/us/crumbley-parents-charged-michigan-shooting.html [https://perma.cc/WGL4-9D69]).

[207] 539 P.2d 36 (Cal. 1975).

[208] *Id.* at 38, 41.

[209] *Id.* at 40.

[210] Bambauer et al., *supra* note 95, at 531.

[211] *Id.* at 530.

[212] *Id.*

[213] *Id.* (first citing 1 JAMES F. MOSHER, LIQUOR LIABILITY LAW § 5.03 (35th ed. 2022); and then citing Peter A. Slepchuk, Note, *Social Host Liability and the Distribution of Alcohol and Narcotics: A Survey and Guide*, 44 SUFFOLK U. L. REV. 933, 933 (2011)).

Masconale, and Simone M. Sepe explained that a human leader can influence others and radicalize them to cause physical harm, and therefore, this is a context that should be subject to liability.[214] Similarly, algorithms that target recommendations are not so different from leaders influencing dynamics on social networks and encouraging users to inflict harm.[215] In fact, such targeting can be even worse, as it harnesses the influence of surveillance capitalism.[216] The more digital companies use surveillance capitalism and technology to influence susceptible end users, the more justified it becomes to hold them responsible for the consequences.[217]

Legislators are already starting to recognize the need to regulate targeting of children.[218] This Article continues this line of thought. Before outlining the proposed framework for imposing a duty of care, the next Part addresses normative free speech considerations for imposing liability on targeting and overviews current case law regarding platform liability.

II. INTERMEDIARY LIABILITY FOR TARGETING: NORMATIVE CONSTITUTIONAL CONSIDERATIONS AND THE LAW

The digital age has created a new model of speech regulation that includes many speech regulators; indeed, the government is no longer the only regulator of individual expression.[219] Owners and operators of the digital infrastructure, through which people "speak" online, also have free speech rights.[220] Some state regulations have attempted to

[214] *See id.* at 506–20, 533–40 (introducing the theory of reckless association as a context of liability).

[215] *Cf. id.* at 491 (suggesting that leaders are influential on social networks because they persistently push certain ideas).

[216] Balkin, *Free Speech*, *supra* note 100, at 1246–49, 1253.

[217] *Id.*

[218] *See, e.g.*, California Age-Appropriate Design Code Act, Assemb. B. 2273, 2021–2022 Reg. Sess. (Cal. 2022) (to be codified in CAL. CIV. CODE § 1798.99.28) (requiring higher privacy protections for online services that are likely to be accessed by children); S. 680, 2023 Leg., Reg. Sess. (Cal. 2023) (imposing liability on social media platforms that do not exercise reasonable care in their design); Natasha Singer, *F.T.C. Seeks 'Blanket' Ban on Meta's Use of Young Users' Data*, N.Y. TIMES (May 3, 2023), https://www.nytimes.com/2023/05/03/technology/facebook-meta-ftc-data-ban-instagram.html [https://perma.cc/KA9R-2N9Y] (discussing how the FTC seeks to expand Meta's legal commitments to improve its privacy for child users). This Article will further address these bills and regulation *infra* Part III.

[219] Balkin, *Free Speech*, *supra* note 100, at 1206.

[220] *Id.*

force digital infrastructure owners to regulate and surveil user speech for governmental objections; this type of government regulation has been dubbed "*new-school speech regulation*."[221] First Amendment doctrine, addressed further below,[222] still has a role in the digital age, yet it continues losing relevance with respect to online algorithmic speech governance, which differs from traditional dissemination of speech.[223] The First Amendment doctrine is thus "inadequate to secure the values that justify it in the first place."[224] Free speech, however, is broader than the First Amendment, and the gap between them widens every day.[225]

A. Normative Analysis: Free Speech in the Algorithmic Society

The algorithmic society model is based on surveillance capitalism and targeting, and under this model, the structure of speech is no longer just a circulation of ideas among autonomous individuals.[226] Instead, it is a business model that collects and measures data and connections to predict social behavior and influence users.[227] This model of speech governance is also a "surveillance system."[228] Further, "much of the decision[-]making is performed by algorithms, not bureaucrats."[229] The system also allows new forms of influence, as it is not just a conduit for free speech; it surveils, analyzes, and shapes the behavior and expression of both the speakers and audiences using the platform.[230] Thus, because this type of model functions through technological affordances and AI agents,[231] the focus is no longer on the circulation of human reason through a system of free speech rights.[232] Indeed, with changes in the structure of speech governance diverging from traditional models, individuals should reassess how they view freedom of expression.[233] In the algorithmic society, there

[221] *Id.* at 1216.

[222] *See infra* Section III.C.3.

[223] Balkin, *Free Speech*, *supra* note 100, at 1213–14.

[224] *Id.* at 1259.

[225] *Id.* at 1273.

[226] *Id.* at 1214.

[227] *Id.* at 1214, 1244.

[228] *Id.* at 1244.

[229] *Id.* at 1241.

[230] *Id.* at 1244.

[231] *Id.* at 1242–43.

[232] *Id.* at 1253 (describing the weaponization of predictive algorithms as a tool to influence, mislead, and misinform end users).

[233] *Id.*

are substantial justifications for revisiting free speech conventions regarding targeting and rethinking how to regulate new forms of informational capitalism "that have had enormous effects on free expression, culture, and politics."[234] Although liability may hamper an intermediary's freedom to conduct business as it sees fit, because targeting recommendations of harmful content to children can result in severe harm, children's safety outweighs the constraints accompanying such liability.[235] The following Sections will examine free speech considerations for imposing liability on intermediaries for targeting children.

1. *Free speech values and the algorithmic society*

One might allege that imposing liability for targeting children with algorithmic recommendations threatens the audience's freedom of speech because holding intermediaries liable could result in collateral censorship, interfering with the user's ability to speak.[236] Indeed, there are several main theories concerning the reason for free speech protections,[237] including protecting individual autonomy and self-fulfillment,[238] as well as facilitating a free marketplace of ideas that allows the pursuit of truth.[239] Another prominent justification for protecting freedom of expression is that it promotes democracy through informed public deliberation of public issues,[240] which is necessary to exercise public power and prevent the abuse of power.[241]

Some may argue that imposing liability on intermediaries could infringe on speakers' autonomy, impair the public's ability to receive information, disrupt the exchange of ideas, and undermine civic and cultural participation.[242] However, the shift from the internet society to the algorithmic society has changed the nature of speech, and therefore, the application of free speech values requires consideration

[234] *Id.* at 1214.

[235] Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 295–96 (2011) [hereinafter Wu, *Collateral Censorship*].

[236] *See id.* at 296–97.

[237] Lavi, *Targeting Exceptions*, *supra* note 77, at 123.

[238] *Id.* at 123 (citing Joseph Raz, *Free Expression and Personal Identification*, 11 OXFORD J. LEGAL STUD. 303, 311–16 (1991)).

[239] Balkin, *Free Speech*, *supra* 100, at 1259; Lavi, *Targeting Exceptions*, *supra* note 77, at 123–24.

[240] Balkin, *Free Speech*, *supra* note 100, at 1259–60; Lavi, *Targeting Exceptions*, *supra* note 77, at 124.

[241] Balkin, *Free Speech*, *supra* note 100, at 1260.

[242] Lavi, *Targeting Exceptions*, *supra* note 77, at 125–26.

of this transition.[243] On the one hand, the digital age greatly increases opportunities for people to express themselves, speak to each other, and reach audiences around the world; on the other hand, it weakens the institutions that secure free speech.[244] Unlike twentieth-century mass media, social media does not produce most of the content;[245] rather, it encourages users to create independent content and for other users to react.[246] This ecology works best by repeating ideas, seeking attention, and generating conflict.[247] Targeting influences the context of the message by controlling the target audience. For example, recommendations and advertisements are purposefully timed and distributed for maximum effect.[248] Targeting recommendations to vulnerable populations, and children in particular, in fact *narrows* exposure to the free marketplace of ideas and hampers the very same values at the foundation of free speech in three distinctive ways:

First, because targeting is personalized and directed at susceptible populations without transparency, it can impair the audience's ability to make informed choices and act autonomously. Children's decision-making abilities are more limited than that of adults[249] and all the more so when their choices are subject to manipulation.[250] Furthermore, targeting content that specifically normalizes violence or self-harm to children chills the target's speech;[251] instead of promoting their autonomy and self-development, targeting undermines it, as personalized targeting of such content manipulates, and even encourages, children to inflict self-harm.[252]

Second, targeting personalized recommendations fails to allow fair competition between ideas in the free marketplace. Instead, it

[243] Balkin, *Free Speech*, *supra* note 100, at 1261–62.

[244] *Id.* (describing the ways digital technologies undermine trust in science, journalism, and other public institutions).

[245] *Id.* at 1265.

[246] *Id.*

[247] *Id.*

[248] *Id.* at 1244, 1258.

[249] On the limitations of children's brains in comparison to adults and the possibility to manipulate them, see Gilad et al., *supra* note 63, at 1295.

[250] *See* Lavi, *Manipulating*, *supra* note 53, at 269–71.

[251] *Id.* at 283.

[252] For a similar argument regarding violent sexual content, see DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY IDENTITY, AND LOVE IN THE DIGITAL AGE 119–25 (2022).

encourages replication of expressions that grab attention.[253] When intermediaries use algorithms to promote biased agendas, they exercise disproportionate power, creating unequal access to information, thereby stifling the marketplace of ideas.[254] Under the targeting model, knowledge is less likely to spread widely, truth is less likely to prevail over falsehood,[255] and content encouraging violence and self-harm, which adds little to the marketplace of ideas, receives more attention, and impairs fair competition between ideas.[256] Thus, targeting can undermine users' free speech and right to receive information.[257]

Third, intermediaries targeting recommendations to children exercise vast influence by leading them to focus on specific types of content at the expense of the diverse content to which they might have otherwise been exposed. In doing so, targeting hinders true discourse and can erode democracy. Targeting also undermines public participation in democratic participatory culture.[258] Algorithmic recommendation systems undermine the public's ability to be informed about public issues because targeting limits the target audience to a narrow set of information.[259] Since individuals have limited attention spans, they might avoid seeking more information beyond the algorithmic recommendations.[260] In addition, targeting often manipulates the target audience by utilizing the information collected on them and reshaping their cultural identities in harmful ways; in short, this could hinder democratic deliberation.[261]

In summary, liability for targeting in the algorithmic society does not infringe on the target audience's free speech; in fact, targeting susceptible individuals prevents users from making meaningful choices about what content to consume and therefore does not promote free speech values. Imposing liability on intermediaries that

[253] Balkin, *Free Speech*, *supra* note 100, at 1265.

[254] Lavi, *Targeting Exceptions*, *supra* note 77, at 133.

[255] Balkin, *Free Speech*, *supra* note 100, at 1262.

[256] On the idea that there are types of content that add little, if at all, to the public sphere, see generally CITRON, *supra* note 252.

[257] Lavi, *Targeting Exceptions*, *supra* note 77, at 133.

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

[259] Balkin, *Free Speech*, *supra* note 100, at 1264.

[260] *Id.*

[261] *Id.* at 1267.

target individuals remedies a distortion the algorithmic society creates, and in turn, promotes the target audience's free speech.

2. *Intermediary free speech in the algorithmic society*

U.S. law requires very minimal elements of "communication" to constitute speech.[262] Accordingly, it could be argued that targeting is considered the intermediary's speech, and thus imposing liability on targeting would infringe upon the intermediary's free speech rights.[263] Under this argument, it could be alleged that recommendations can extend well beyond a functional tool: "[T]he tool itself is an expression of the intermediary's ideas or serves as advice to users."[264]

However, targeting in the algorithmic society is based on an analysis of collected information carried out by AI algorithms, not humans.[265] Many scholars argue that AI algorithms should not be assigned the same scope of free speech rights as humans and should instead be assigned only secondary speech rights, if any.[266] Other scholars even assert that AI outputs are not speech, as algorithms do not intend to

[262] Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1461 (2013); Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 1004 (2019).

[263] For further information on why targeting is a form of protected commercial speech, see Lavi, *Targeting Exceptions*, *supra* note 77, at 137.

[264] Lavi, *Do Platforms Kill?*, *supra* note 44, at 531–32 (citing James Grimmelmann, *Speech Engines*, 98 MINN. L. REV. 868, 874 (2014)); *see* Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1515–21 (2013) [hereinafter Wu, *Machine Speech*].

[265] *See, e.g.*, Benjamin, *supra* note 262, at 1464–66.

[266] Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 46; Nathan Cortez & William M. Sage, *The Disembodied First Amendment*, 100 WASH U. L. REV. 707, 710 (2023) ("[C]ourts return to the original justification for covering commercial speech—protecting the interests of human consumers and listeners—and abandon later justifications that look to the interests of non-human speakers or the value of information for its own sake."); FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTS IN THE AGE OF AI 109 (2020) ("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, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY 273, 281 (Lee Bollinger & Geoffrey Stone eds., 2022) ("[T]he replicant targeting the ads in Facebook's algorithm would have no presumptive constitutional protection.").

communicate[267] since they are only tools aimed at assisting users in finding content that interests them.[268]

By personalizing recommendations and targeting them to listeners, algorithms no longer act as a passive conduit, but they autonomously add their own elements;[269] accordingly, AI outputs are non-speech and are therefore not constitutionally protected.[270] The algorithm does not "know" what the recommended content is or what the user data signifies;[271] rather, "[i]t only knows correlations, which it calculates mathematically."[272]

Even according to scholars who recognize machine speech as protected speech, the protection of such speech is not at the core of the free speech justifications, nor is it even at the core of First Amendment protection. Accordingly, even if algorithmically-generated content were viewed as speech, and even if that speech was imbued with some constitutional protection against government

[267] *See* Toni M. Massaro, Helen Norton & Margot E. Kaminski, *SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481, 2507 (2017) (questioning how a court would assess liability when culpable intent cannot be shown).

[268] Lavi, *Do Platforms Kill?*, *supra* note 44, at 531; Wu, *Machine Speech*, *supra* note 264, at 1504; *see also* Massaro et al., *supra* note 267, at 2483–84 (suggesting current First Amendment jurisprudence constrains the extent to which AI can be regulated as speech).

[269] *See* Dr. Lavi Brief of Amicus Curiae, *supra* note 16, at 15 (using the Netflix series *13 Reasons Why* to exemplify how algorithms change the context of the content through targeting).

[270] KARL M. MANHEIM & JEFFERY ATIK, WHITE PAPER: AI OUTPUTS AND THE FIRST AMENDMENT 3 (2023) ("Most AI, including social media recommendation algorithms and LLMs, are at least semi-autonomous in that they design their own outputs and do not merely restate human inputs in contextually relevant ways. Their communication can be in the form of human language or other human expression (e.g., graphics, audio, video), which makes it easy to mistake it for speech. It is not, but if it were, whose speech would it be? It is not constructed by a human, unless the programming is so specific and granular that the machine is no longer acting as an artificially intelligent agent. Rather, an AI's output is fully composed by the machine itself, thus lacking the factors that convert communication into speech."); Dan L. Burk, *Asemic Defamation, or, The Death of the AI Speaker*, 22 FIRST AMEND. L. REV. (forthcoming 2024) ("[F]or LLM-generated texts there is no speaker—and hence no communicative meaning—at all.").

[271] MANHEIM & ATIK, *supra* note 270, at 4.

[272] *Id.*

regulation, given children's unique susceptibilities, such speech should still be regulated.[273]

Some experts allege that even if algorithm owners do not claim free speech rights for machine speech as speakers, the public's right to information could justify the protection of machine speech because it can serve the listener's free speech interests.[274] Even still, this "listener-centered" approach would allow the regulation of speech made by knowledgeable or powerful speakers if that speech conflicts with the listener's autonomy.[275] Intermediaries collect data on users and use it for algorithmic recommendations,[276] making them powerful, knowledgeable speakers and justifying the application of a "listener-centered" approach for government regulation.[277] Intermediary recommendations manipulate, at the very least, their young targets and frustrate their autonomy and self-governance, reshaping their cultural identities in harmful ways.[278] Accordingly, intermediaries do not promote the free speech of such listeners, and therefore, the regulation of such targeting is justifiable.[279] Hence, machine speech should not benefit from overall protection, especially when targeting can result in enormous harm to children.[280]

Another more far-reaching argument is that the listener's right to receive information should be limited to actual speech-human

[273] Inyoung Cheong, *Freedom of Algorithmic Expression*, 91 U. CIN. L. REV. 680, 682–83 (2023); Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 46.

[274] Helen Norton, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441, 443, 451 (2019) [hereinafter Norton, *Powerful Speakers*]; Helen Norton, *Manipulation and the First Amendment*, 30 WM. & MARY BILL RTS. J. 221, 230–31 (2021) [hereinafter Norton, *Manipulation*]; MANHEIM & ATIK, *supra* note 270, at 4 ("Algorithmic outputs in any particular instance may reflect the desires of the targeted 'listeners' more than the intention of any sender, autonomous or human.").

[275] Norton, *Powerful Speakers*, *supra* note 274, at 441–42; Lavi, *Manipulating*, *supra* note 53, at 289–90.

[276] *See* Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, 1 J. FREE SPEECH L. 71, 84 (2021) [hereinafter Balkin, *Regulating Social Media*] (explaining how massive technology companies use all the data they collect to train algorithms and better predict user behavior).

[277] Norton, *Powerful Speakers*, *supra* note 274, at 441–42.

[278] *Id.*

[279] *Id.*

[280] *See* Cortez & Sage, *supra* note 266, at 710, 711, 760 (arguing that the "balance between public regulation and individual liberty must account for the harms caused by modern forms of corporate and artificial speech").

communication.[281] In other words, purely AI-generated information should not be considered speech at all and should not be protected by the First Amendment.[282] Commentators have suggested that AI speech should not receive protection because it does not contribute to the marketplace of ideas since AI does not have any ideas of its own, regardless of the ideas it may spawn in its viewers.[283] Yet even if machine-generated communications inspire listeners, it is doubtful whether this machine speech could receive First Amendment protection under a right to receive information theory.[284]

3. Does liability for targeting children chill speech?

Arguably, imposing liability on targeting could result in extensive collateral censorship of recommendations and reduce efficiency and innovation.[285] However, even if imposing liability in these cases could result in over-censorship of legitimate recommendations targeted to adults, this would not even affect the whole public.[286] When intermediaries self-censor recommendations, it should be treated differently than censoring users' speech because these recommendations are derived from third-party content, they are not content themselves.[287] Recommendations direct users to content they do not specifically seek;[288] when liability is directed at the intermediary's own recommendations, users remain free to seek out the content themselves.[289]

Moreover, extending liability to targeting children would not lead to disproportionate collateral censorship. Indeed, websites would

[281] MANHEIM & ATIK, *supra* note 270, at 6.

[282] *Id.* at 5.

[283] *Id.* at 7.

[284] *Id.; see also* Burk, *supra* note 270, at 26 ("Constitutional protection of listener interests inherently assumes the presence of an intentional speaker. The 'listener' may find meaning in the asemic text, but it is never the meaning intended by the text's author, because there is none.").

[285] Wu, *Collateral Censorship*, *supra* note 236, at 295–96 ("Collateral censorship occurs when a (private) intermediary suppresses the speech of others in order to avoid liability that otherwise might be imposed on it as a result of that speech.").

[286] *See* Smith v. California, 361 U.S. 147, 153–54 (1959) (striking down an ordinance imposing liability on sellers of obscene books on grounds that the fear of liability would result in an over-correction of self-censorship that would affect the whole public).

[287] Lavi, *Do Platforms Kill?*, *supra* note 44, at 531.

[288] Lavi, *Targeting Exceptions*, *supra* note 77, at 132.

[289] Lavi, *Do Platforms Kill?*, *supra* note 44, at 531.

continue to operate and could target other populations, and because intermediaries would continue to earn profit from targeting adult audiences, they would still have an incentive to operate recommendation systems.[290]

B. The Law: § 230 of the Communications Decency Act and Liability

Currently, a debate exists over the extent to which § 230 of the Communications Decency Act (CDA)[291] "bars regulation of website conduct to prioritize, recommend, or otherwise steer users toward particular content."[292] Section 230 of the CDA provides immunity to internet users and intermediaries that disseminate information created by others.[293] U.S. courts have addressed liability for negligent design[294] and targeting in particular; however, many courts have interpreted § 230 broadly and have avoided addressing substantive questions of liability.[295] Some courts have recognized exceptions to the immunity regime; yet, even where immunity is not applied, court decisions reveal substantive difficulties and often avoid imposing liability. The following Section will address court decisions regarding intermediary liability and legislative proposals to narrow § 230 with respect to algorithmic targeting.

Section 230(c)(1) of the CDA 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

[290] *See* Balkin, *Regulating Social Media*, *supra* note 276, at 94 (explaining how establishing distributor liability for paid advertisements on social media platforms might mitigate the problems of collateral censorship).

[291] 47 U.S.C. § 230.

[292] Lawrence, *supra* note 83, at 14–15.

[293] *See, e.g.*, Barrett v. Rosenthal, 146 P.3d 510, 519, 528–29 (Cal. 2006) (observing that the plain language of § 230 is evidence that Congress did not intend for an internet user to be treated differently than an internet provider). *See generally* Lawrence, *supra* note 83 (noting that § 230 is a "pivotal" limitation on states' authority to regulate online content).

[294] A negligent design claim arises when a product's design is flawed or unsafe, causing harm to consumers or users. *See, e.g.*, Lemmon v. Snap, Inc., 995 F.3d 1085, 1091–94 (9th Cir. 2021) (discussing negligent design claims and upholding a claim against Snapchat for negligently designing its platform to include a filter that encouraged users to drive recklessly).

[295] *See infra* notes 317–29 and accompanying text (providing an overview of cases defining § 230 broadly.

provider."[296] Under subsection (c) titled "Protection for 'Good Samaritan' Blocking and Screening of Offensive Material," Congress declared that online intermediaries can never be treated as "publishers" of material they did not develop.[297] The Section differentiates between the internet and the media that preceded it, "represent[ing] 'the mindset of internet exceptionalism'" by generally blocking lawsuits against online intermediaries.[298] In passing § 230, Congress aimed to promote self-regulation and free speech, fostering the rise of vibrant internet enterprises.[299]

Courts have reflected the strong U.S. bias in favor of free speech and its presumption against speech restrictions by interpreting § 230 broadly and blocking lawsuits against intermediaries.[300] In *Zeran v. America Online, Inc.*,[301] for example, the Fourth Circuit held that AOL, the intermediary, negligently failed to remove an anonymous online bulletin board post including the plaintiff's phone number and claiming the plaintiff was selling T-shirts glorifying the Oklahoma City bombing.[302] Consequently, the plaintiff's phone rang incessantly, and he even received death threats.[303] The Fourth Circuit held that the distributors were immune from liability because they were a subset of publishers.[304] According to *Zeran*, § 230 provides site hosts immunity regardless of whether the hosts act on knowledge of illegal content on

[296] 47 U.S.C. § 230(c)(1) (emphasis added); *see* Lawrence, *supra* note 83, at 14 (suggesting that this Section also may protect platforms from liability even if they fail to censor content). For an overview of the history and power of § 230, see CITRON, *supra* note 252, at 84–86 and United States—Mexico—Canada Agreement art. 19.17, Nov. 30, 2018, OFF. U.S. TRADE REPRESENTATIVE [hereinafter USMCA], https://ustr.gov/trade-agreements/free-trade-agreements/united-states-mexico-canada-agreement/agreement-between [https://perma.cc/PLR7-9MVQ] (exporting § 230 to Canada and Mexico through the United States-Mexico-Canada Trade Agreement, which went into effect on July 1, 2020).

[297] *See* 47 U.S.C. § 230(c).

[298] Hadar Y. Jabotinsky & Michal Lavi, *NFT for Eternity*, 56 U. MICH. J. L. REFORM 827, 869 (2023) (quoting Lavi, *Re-Tweet*, *supra* note 110, at 486).

[299] Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639, 652 (2014).

[300] Lavi, *Targeting Exceptions*, *supra* note 77, at 103.

[301] 129 F.3d 327 (4th Cir. 1997).

[302] *Id.* at 329.

[303] *Id.*

[304] *See id.* at 328, 332 (explaining that although the distinctions between distributors and publishers come from defamation law, AOL fit the legal definition of a publisher, so it could not be held not liable).

their sites or fail to act at all.[305] Following *Zeran*, courts have repeatedly used § 230 to shield intermediaries and other web enterprises from liability.[306]

Courts have upheld immunity even when the intermediary's role extended beyond content moderation. In *Blumenthal v. Drudge*,[307] for example, the court granted immunity to the intermediary, AOL, after AOL paid an independent contractor to write gossip columns containing defamation for the site.[308] Similarly, in *Batzel v. Smith*,[309] a website operator and electronic listserv for Museum Security Network ("MSN") received an email with false accusations that Ellen Batzel had hundreds of old European paintings, likely looted during World War II, hanging on her walls.[310] The MSN operator edited and then publicly posted the defamatory email on the network and website even though the sender did not intend for the email to be posted.[311] Batzel sued the listserv's editor.[312] The Ninth Circuit concluded that the MSN operator was immune from liability because, even though the listserv editor could exercise some control over the listserv messages, the listserv was an interactive computer service provider under § 230.[313] Judge Gould dissented, concluding that the listserv editor had constructed content worthy of dissemination by selecting and publishing material not intended to be published online.[314] The majority, however, applied immunity even when the intermediary's role exceeded mere hosting and included publicly publishing information sent privately by a third party.[315]

[305] *See id.* at 328 (holding that AOL was immunized under § 230 even for its failure to act).

[306] Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 402, 406 (2017); Lavi, *Targeting Exceptions*, *supra* note 77, at 104; *see, e.g.*, Klayman v. Zuckerberg, 753 F.3d 1354, 1355 (D.C. Cir. 2014) (holding that § 230 protected Facebook from the plaintiff's action after Facebook allegedly delayed removing violent and offensive content from its site); Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 18, 22 (1st Cir. 2016) (relying on the language of § 230 and the holding in *Zeran* to find the website free from liability).

[307] 992 F. Supp. 44 (D.D.C. 1998).

[308] *Id.* at 51, 53; Lavi, *Targeting Exceptions*, *supra* note 77, at 104–05.

[309] 333 F.3d 1018 (9th Cir. 2003).

[310] *Id.* at 1021.

[311] *Id.* at 1022.

[312] *Id.*

[313] *Id.* at 1031.

[314] *Id.* at 1038 (Gould, J., dissenting).

[315] *Id.* at 1034 (majority opinion).

1. *Immunity, exceptions, and inconsistency regarding immunity and substantive questions of liability*

Although courts have generally applied immunity broadly, there has been a gradual erosion of overall immunity, and some courts have refrained from applying it altogether. Thus far, courts have based denial of § 230 immunity on two primary arguments: "(1) where the platform at least partly developed or created the content; and (2) where the claim did not treat the platform as the publisher or speaker of third-party content."[316] Yet, courts have been inconsistent in denying § 230 immunity even in these cases, particularly regarding platform design and algorithmic recommendations and targeting.[317]

a. Developing content and the gradual erosion of overall immunity: Fair Housing Council v. Roommates.com

Fair Housing Council v. Roommates.com, LLC,[318] involved a popular roommate-matching website that helps users find roommates.[319] The website requires users to create a personal profile and answer questions regarding their gender, sexual orientation, and parental status, as well as to express their preferences of roommates on each of these topics.[320] Users selected answers from drop-down menus and used an internal search engine, which provided filters based on those questions, to find roommates while filtering out unsuitable matches.[321] The site sent users emails with potential roommate matches from time to time.[322] The Fair Housing Council ("FHC") sued Roommates.com, alleging that the drop-down menu questions, the internal search engine, the

[316] Jeff Kosseff, *A User's Guide to Section 230, and a Legislator's Guide to Amending It (or Not)*, 37 BERKELEY TECH. L.J. 757, 779–80 (2022); Lavi, *Targeting Exceptions*, *supra* note 77, at 110. These exceptions appear in the text of § 230 itself: an information content provider will not be "treated as the publisher or speaker." 47 U.S.C. § 230(c)(1). Congress defined an "information content provider" as "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the [i]nternet or any other interactive computer service." § 230(f)(3).

[317] *See* Lavi, *Targeting Exceptions*, *supra* note 77, at 118, 171 (explaining that while courts have generally trended toward finding that websites are not information providers, immunity under § 230 is less certain than it used to be because some courts have challenged the traditional view of the statute).

[318] 521 F.3d 1157 (9th Cir. 2008) (en banc).

[319] *Id.* at 1161.

[320] *Id.*

[321] *Id.* at 1165.

[322] *Id.* at 1162.

filtering service, and the open comment section were discriminatory and violated the Fair Housing Act.[323] According to the FHC, by conditioning participation in the service upon reporting restricted information, Roommates.com functioned as an information content developer.[324]

On appeal, the Ninth Circuit declined to apply immunity.[325] Chief Justice Kozinski concluded for the majority that the site was a developer rather than a mere "passive transmitter" of information because its questionnaire contained preidentified answer choices.[326] The court reasoned that because the site distributed discriminatory content via its internal search engine and email mechanism, it was not entitled to immunity for its search engine and email.[327] The court further invoked the 'material contribution to illegality' test, which denies immunity when a defendant's own actions materially contribute to the illegality,[328] and concluded that using "*neutral* tools to carry out what may be unlawful or illicit searches does not amount to 'development' for [Section 230] immunity . . . ."[329] The drop-down menus, however, fostered illegal, discriminatory content; therefore, Roommates.com was held liable for the discriminatory content.[330] The court upheld immunity for content posted in the open comment section.[331]

b. Gradual erosion: negligent design—Not a publisher or speaker

Recent cases regarding negligent design of applications have applied *Roommates.com* and have not applied § 230 immunity to digital product design.[332] Court decisions have even overcome the difficulty of demonstrating causal causation beyond § 230, addressing questions of substantive liability, and recognizing that a design can create a context of susceptibility.

[323] *Id.*

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

[325] *Id.* at 1175.

[326] *Id.* at 1166.

[327] *Id.* at 1167.

[328] *Id.* at 1168; Lavi, *Targeting Exceptions*, *supra* note 77, at 112.

[329] *Roommates.com*, 521 F.3d at 1169.

[330] *Id.* at 1167.

[331] *Id.* at 1174.

[332] *See* 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, 22 (2016) (finding that fourteen of twenty-seven opinions ruling on § 230 immunity between 2015 and 2016 declined to provide full immunity).

In *Lemmon v. Snap, Inc.*,[333] for example, two boys died in a high-speed car accident.[334] Their parents sued Snapchat's application provider, Snap, Inc., alleging that the negligent design of Snapchat's Speed Filter encouraged the boys to drive recklessly, resulting in the fatal car crash.[335] The Ninth Circuit reversed the district court's dismissal of the plaintiffs' amended complaint alleging Snap, Inc. did not enjoy § 230 immunity, thereby allowing the plaintiffs' negligent design lawsuit to proceed.[336] Pointing to its decision in *Roommates.com*, the court focused its decision on Snapchat's design and not on whether Snap, Inc., was a publisher or speaker.[337] Therefore, the court refrained from applying § 230 immunity[338] and found that Snap, Inc., could be held liable for unreasonably negligent design of the filter.[339] On remand from the Ninth Circuit's § 230 denial, the U.S. District Court for the Central District of California held:

> [T]he Speed Filter's design encouraged Plaintiffs to drive at dangerous speeds. If Snapchat users are seeking to obtain an unknown trophy associated with using the Speed Filter, it is plausible that they would seek this trophy by increasing their speed—the only metric recorded by the Speed Filter. Even if there were no reward system whatsoever, the basic design of the Speed Filter itself appears to encourage reckless driving. There is realistically no purpose for the Speed Filter other than to encourage users to travel at high speeds and record themselves doing so.[340]

[333] 995 F.3d 1085 (9th Cir. 2021).

[334] *Id.* at 1087.

[335] *Id.* at 1087–88.

[336] *Id.* at 1087.

[337] *See id.* at 1091, 1093 (citing Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1162 (9th Cir. 2008) (en banc)) (applying the *Barnes* test to reject liability in cases where plaintiffs treat the internet company in question as a products manufacturer rather than a publisher or speaker).

[338] Tyler Lisea, Lemmon *Leads the Way to Algorithm Liability: Navigating the Internet Immunity Labyrinth*, 50 PEPP. L. REV. 785, 806 (2023) ("The court deemed Snap's duty to design a reasonably safe product to be fully independent of its role in monitoring and publishing third-party content.").

[339] *Lemmon*, 995 F.3d at 1094.

[340] Lemmon v. Snap, Inc., No. CV 19-4504-MWF, 2022 WL 1407936, at *7 (C.D. Cal. Mar. 31, 2022).

The court also addressed the substantive question of causal connection, holding that a strong causal connection existed between the Speed Filter and the speeding:[341]

> [T]here is no 'gap' between the design of the Speed Filter and the Plaintiffs' accident. . . . It is *extremely* foreseeable that minors and young adults would use the Speed Filter to record themselves driving at excessive speeds, and even more so if there are potential reward 'trophies' for so doing.[342]

In *Maynard v. Snapchat, Inc.*,[343] the court recognized a duty of care and imposed liability for the same Speed Filter.[344] Christal McGee drove three passengers in her family's car at an excessive speed, attempting to reach 100 miles per hour to capture the speed in a photo using Snapchat's Speed Filter.[345] Her car hit Maynard's car, causing permanent brain damage to a passenger in Maynard's car.[346] Maynard sued Snapchat, arguing that Snapchat knew its users might use its service in a manner that would distract them from obeying traffic laws and encouraged dangerous speeding, resulting in the accident.[347] The lower court found that § 230 immunity applied and dismissed the case.[348] Maynard then appealed, arguing that the Complaint contained plausible allegations that Snap breached its duty to exercise reasonable care.[349] The Georgia Court of Appeals affirmed the dismissal, holding that Snap did not have a duty under tort law to adjust the Speed Filter to prevent intentional misuse by third parties.[350] However, the Georgia Supreme Court reversed the Court of Appeal's judgment, accepting

[341] *See id.* at *10 ("[G]iven that the accident occurred while the Plaintiffs were using the Speed Filter for the exact purpose for which it appears to have been designed: to record the user traveling at excessive speeds.").

[342] *Id.*

[343] 870 S.E.2d 739 (Ga. 2022), *remanded to* 883 S.E.2d 533 (Ga. App. 2023).

[344] *Id.* at 743.

[345] *Id.*

[346] Eric Goldman, *Snapchat May Have a Duty Not to Design Dangerous Software*—Maynard v. Snap, TECH. & MKTG. L. BLOG (Mar. 18, 2022) [hereinafter Goldman, *Snapchat May Have a Duty*], https://blog.ericgoldman.org/archives/2022/03/snapchat-may-have-a-duty-not-to-design-dangerous-software-maynard-v-snap.htm [https://perma.cc/K3EC-YH2W].

[347] *Maynard*, 870 S.E.2d at 744 ("Speed Filter was motivating, incentivizing, or otherwise encouraging its users to drive at excessive, dangerous speeds in violation of traffic and safety laws.").

[348] Goldman, *Snapchat May Have a Duty*, *supra* note 346.

[349] Maynard v. Snapchat, Inc., 851 S.E.2d 128, 130 (Ga. 2020), *rev'd*, 870 S.E.2d 739 (Ga. 2022).

[350] *Id.* at 133.

the argument "that Snap could reasonably foresee the particular risk of harm from the Speed Filter."[351] The case was remanded for further proceedings.[352]

These cases pave the way to hold companies liable for their algorithmic designs.[353] Nevertheless, courts have issued conflicting judicial decisions on the scope of immunity with respect to algorithmic recommendations and targeting. As demonstrated by the Sections below, many decisions upheld immunity.

c. Algorithmic design: Matching, recommending, and targeting

In *Dyroff v. Ultimate Software Group, Inc.*,[354] Ultimate Software used machine learning algorithms to assess users' intent and emotional state, aiming "to steer users to particular groups."[355] Its notification and recommendation functions included "push notifications, which alerted users of new content posted to its groups."[356] During a Google search for ways to purchase heroin, Wesley Greer was directed to a group on the platform titled "where can i score heroin in jacksonville, fl. [sic]."[357] Greer posted to the group, and soon after, he received an email notification from the platform that another user, Hugo Margenat-Castro, had responded with a hyperlink.[358] Greer and Castro communicated via the platform, Castro sold Greer heroin laced with fentanyl, and Greer "died from fentanyl toxicity, unaware of its presence in the heroin."[359] Kristanalea Dyroff, Greer's mother, filed a suit against Ultimate Software.[360] The Ninth Circuit held that § 230 of the CDA immunized the intermediary and concluded that by recommending user groups and sending email notifications, Ultimate Software acted as a publisher of others' content.[361] According to the court, while recommendations and notifications help users communicate with one another, these functions did not meaningfully

[351] *Maynard*, 870 S.E.2d at 743.

[352] *Id.*

[353] *See* Lisea, *supra* note 338, at 806.

[354] No. 17-cv-05359-LB, 2017 WL 5665670 (N.D. Cal. Nov. 26, 2017), *aff'd*, 934 F.3d 1093 (9th Cir. 2019).

[355] Rifkind, *supra* note 67, at 54.

[356] *Id.*; *Dyroff*, 934 F.3d at 1098.

[357] Rifkind, *supra* note 67, at 54.

[358] *Id.*

[359] *Id.*

[360] *Id.* at 55.

[361] *Id.* at 55, 56.

contribute to the illegal content.[362] The U.S. Supreme Court denied Dyroff's petition for certiorari.[363]

In *Daniel v. Armslist, LLC*,[364] Armslist.com allowed buyers and sellers of firearms to contact each other.[365] The design of the platform permitted sales of illegal firearms, one of which was eventually used in a deadly shooting.[366] The plaintiff claimed that the design and operational characteristics of Armslist.com affirmatively "encouraged" the illegal purchase of firearms.[367] The court broadly construed *Roommates.com* and declined to extend immunity to website design features despite the fact that some sales were legal.[368] However, the Wisconsin Supreme Court overturned the lower court's decision, reasoning that Armslist provided *neutral* tools that could be used legally.[369] The court also clarified "that Armslist was not an information content provider, [thereby] dismissing all of the plaintiff's claims."[370]

In *Anderson v. TikTok, Inc.*,[371] Taiwanna Anderson sued TikTok after her eleven-year-old daughter, Nylah, died attempting a dangerous challenge she saw on TikTok.[372] According to Anderson, TikTok knew its algorithm promoted the "Blackout Challenge" to children, yet

[362] Lavi, *Do Platforms Kill? supra* note 44, at 515–16; Lavi, *Targeting Exceptions*, *supra* note 77, at 116.

[363] Dyroff v. Ultimate Software Grp., Inc., 34 F.3d 1093 (9th Cir. 2019), cert. denied, 140 S. Ct. 2761 (2020).

[364] 913 N.W.2d 211 (Wis. Ct. App. 2018), *rev'd*, 926 N.W.2d 710, 714 (Wis. 2019).

[365] *Id.* at 215.

[366] *Id.* at 213–14.

[367] *Id.*

[368] *Id.* at 223–24; Lavi, *Targeting Exceptions*, *supra* note 77, at 116; *Daniel*, 913 N.W.2d at 213–14.

[369] Lavi, *Targeting Exceptions*, *supra* note 77, at 117.

[370] Michael L. Rustad & Thomas H. Koenig, *The Case for a CDA Section 230 Notice-and-Takedown Duty*, 23 NEV. L.J. 533, 555 (2023); *see* Daniel v. Armslist, LLC, 926 N.W.2d 710, 726 (Wis. 2019) (holding that the conduct of publishers like Armslist is protected by § 230); *see also* Lavi, *Do Platforms Kill?*, *supra* note 44, at 516 (explaining that immunity applies regardless of Armslist's knowledge because "§ 230 does not contain a good faith requirement"). *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, 52 (2020) (criticizing the Wisconsin Supreme Court's expansion of § 230 immunity because Armslist does not offer services related to speech).

[371] 637 F. Supp. 3d 276 (E.D. Pa. 2022).

[372] *Id.* at 278; *see supra* notes 35–41 and accompanying text for Nylah's story.

continued to let the algorithm do so.[373] Anderson claimed that TikTok's algorithm

> recommend[s] inappropriate, *dangerous, and deadly videos* to users in a ['For You Page' ("FYP") that] are designed to . . . manipulate them into participating in *dangerous and deadly challenges*; [Tik-Tok's algorithms are] not equipped, programmed with, or developed with the necessary safeguards required to prevent circulation of *dangerous and deadly videos*; and [f]ail[] to warn users of the risks associated *with dangerous and deadly videos and challenges*.[374]

The court dismissed the case based on § 230, explaining that although the algorithmic recommendations are deliberate actions, "such algorithms are 'not content in and of themselves.'"[375] Thus, Anderson's claims were barred under § 230;[376] the court concluded that TikTok was entitled to § 230 immunity because Anderson's claims were "inextricably linked" to TikTok's conduct as a publisher.[377]

> Nylah Anderson's death was caused by her attempt to take up the 'Blackout Challenge.' Defendants did not create the Challenge; rather, they made it readily available on their site. Defendants' algorithm was a way to bring the Challenge to the attention of those likely to be most interested in it. In thus promoting the work of others, Defendants published that work—exactly the activity Section 230 shields from liability.[378]

The court thereby interpreted § 230 broadly and dismissed the case.[379]

[373] *Anderson*, 637 F. Supp. 3d at 278; Eric Goldman, *Section 230 Protects TikTok for "Blackout Challenge" Death, Despite the Algorithms*—Anderson v. TikTok, TECH. & MKTG. L. BLOG (Oct. 27, 2022) [hereinafter Goldman, *Section 230*], https://blog.ericgoldman.org/archives/2022/10/section-230-protects-tiktok-for-blackout-challenge-death-despite-the-algorithms-anderson-v-tiktok.htm [https://perma.cc/YYF2-XK8X].

[374] *Anderson*, 637 F. Supp. 3d at 278 (internal quotations omitted).

[375] *Id.* at 280 (citing Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093, 1098 (9th Cir. 2019)).

[376] *Id.* at 281; *see* Goldman, *Section 230*, *supra* note 373 (explaining that although Anderson's claim was not directed at TikTok as a publisher, publishers are more than hosts of users' content because they promote and distribute content).

[377] *Anderson*, 637 F. Supp. 3d at 281.

[378] *Id.* at 282.

[379] *Id.* at 278.

d. Algorithmic targeting and anti-terrorism statutes cases: § 230 and beyond—The statute requirements for proximate cause, scienter, and substantial assistance

A series of court decisions addressed questions of immunity for platform design and liability for algorithmic recommendations that incite terrorism under § 2333(a) of the Anti-Terrorism Act.[380] Even when terrorism was involved, most courts applied immunity broadly and granted motions to dismiss based on § 230.[381] The courts rejected suits even when plaintiffs based their claims on intermediaries' direct liability and involvement in creating or developing information by targeting messages.[382]

Even when immunity was not applied, a handful of cases focused on the presence of terror organizations on social media and were not dismissed on the basis of § 230.[383] Rather, they were dismissed for lacking elements required to impose liability in Anti-Terrorism statutes, in particular, the requirement for proximate cause; in other words, they were dismissed because the plaintiff could not show sufficient causal connection between the intermediary's personalized recommendations of content and the terror attack.[384]

In *Force v. Facebook*,[385] the plaintiffs, family members of victims of Hamas terrorist attacks in Israel, filed a lawsuit against Facebook, arguing that it materially supported terrorism.[386] The plaintiffs claimed Facebook substantially assisted and "serve[d] as a broker" and network between terrorists.[387] The district court dismissed the case based on § 230.[388] In the appeal, plaintiffs argued that "providing a forum for

[380] 18 U.S.C. § 2333(a) (2018); *see* Lavi, *Do Platforms Kill?*, *supra* note 44, at 517 (noting a causal connection is required because it falls under tort law).

[381] Lavi, *Do Platforms Kill?*, *supra* note 44, at 517.

[382] *Id.*

[383] *Id.* at 521–25.

[384] *Id.*; Fields v. Twitter, Inc., 881 F.3d 739, 741 (9th Cir. 2018); Pennie v. Twitter, Inc., 281 F. Supp. 3d 874, 876 (N.D. Cal. 2017); Sinclair v. Twitter, Inc., No. C 17-5710 SBA, 2019 WL 10252752, at *3–4 (N.D. Cal. Mar. 20, 2019); *see* Cain v. Twitter, Inc., No. 17-cv-02506-JD, 2018 WL 4657275, at *2 (N.D. Cal. Sept. 24, 2018) (dismissing the case because of lack of proximate cause); *see also* Crosby v. Twitter, Inc., 921 F.3d 617, 624–26 (6th Cir. 2019) (rejecting the material support claim because of the proximate cause requirements).

[385] 934 F.3d 53 (2d Cir. 2019).

[386] *Id.* at 57.

[387] *See* Complaint at 2, *Force*, 934 F.3d 53 (No. 18-397).

[388] Cohen v. Facebook, 252 F. Supp. 3d 145, 158–61 (E.D.N.Y. 2017), *aff'd in part sub nom. Force*, 934 F.3d at 53.

communication for terrorists, facilitating personalized 'newsfeed' pages for each user, and providing 'friends suggestions' by using algorithms extend beyond a function of an information content provider" and that Facebook functioned as a *creator* of the information, through algorithmic targeting.[389] The Second Circuit, however, affirmed the district court's opinion holding Facebook immune under § 230 for civilian anti-terrorism claims.[390] The majority held that personalized algorithmic recommendations that impact what content users see in their newsfeeds did not render Facebook a creator or developer of the content because Facebook's algorithms are content-neutral and merely display other users' content to users.[391] The court concluded that making Hamas' content more visible, available, and usable through algorithms did not amount to developing content.[392]

Chief Judge Katzmann dissented from the majority's conclusion on the issue of immunity[393] and declined to apply immunity regarding algorithmic recommendations, reasoning that the claim was not based on the content of information shown but rather on the connections algorithms made between individuals.[394] Thus, Judge Katzmann based his opinion on the argument of algorithmic design. Force filed a petition for writ of certiorari with the U.S. Supreme Court, which declined to hear the case.[395]

The question of liability for algorithmic recommendation ultimately reached the U.S. Supreme Court in *Gonzalez v. Google, LLC*, which concerned coordinated attacks by ISIS on November 13, 2015, at the La Belle Bistro, the Stade de France, and the Bataclan Theater in France.[396] Nohemi Gonzalez, an American citizen, was killed during the attack at La Belle Bistro.[397] Gonzalez's family filed suit against Google (the owner of YouTube), Facebook, and Twitter, alleging the

[389] *Force*, 934 F.3d at 58–59, 64–65.

[390] *Id.* at 66, 68, 72; Cyphert & Martin, *supra* note 43, at 174.

[391] *Force*, 934 F.3d at 68, 70.

[392] *Id.* at 70.

[393] *Id.* at 76–77.

[394] *Id.* For further information, see Lawrence, *supra* note 83, at 15.

[395] Petition for a Writ of Certiorari, *Force*, 934 F.3d 53 (No. 19-859); Adi Robertson, *Supreme Court Rejects Lawsuit Against Facebook for Hosting Terrorists*, VERGE (May 18, 2020, 11:38 AM), https://www.theverge.com/2020/5/18/21262248/supreme-court-rejects-stuart-force-facebook-section-230-lawsuit-algorithms [https://perma.cc/Y6SC-YG7E].

[396] Gonzalez v. Google, Inc., 282 F. Supp. 3d 1150, 1154 (N.D. Cal. 2017), *vacated sub nom.* Gonzalez v. Google LLC, 598 U.S. 617 (2023) (per curiam), *remanded to* 71 F.4th 1200 (9th Cir. 2023).

[397] *Id.*

defendants "knowingly permitted ISIS to use their social networks as a tool for spreading extremist propaganda, raising funds, and attracting new recruits in violation of the [sic] § 2333."[398] The plaintiffs contended that the intermediaries' use of algorithms promoted terrorist propaganda and even directed users to similar videos and accounts, some of which belonged to ISIS.[399] Moreover, the plaintiffs alleged that Google derived income by providing targeted advertisements based on algorithmic analyses of user data.[400] The plaintiffs pointed to *Roommates.com* for support.[401] At a later hearing, plaintiffs sought to distinguish the facts from *Roommates.com* and suggested "that the *Roommates.com* material contribution test" does not apply because the *Roommates.com* court was not considering the exact question and did not address whether targeted advertising is a form of content development.[402] The court, however, dismissed the case, concluding that responsibility for creation or development of content when a website operator materially contributed to unlawfulness is not limited to the facts of the *Roommates.com* case.[403] However, "Plaintiffs do not allege that Google's own actions—here, its targeted ad algorithm—contribute in any way to what makes the ISIS-related videos unlawful or objectionable,"[404]and because "Google's ad pairings do nothing to enhance the unlawfulness of ISIS videos,"[405] there is no material contribution to ISIS's actual videos and immunity applied.[406] Google used "neutral tools" in targeting advertisements and, therefore, did not develop unlawful content.[407]

On appeal, the Ninth Circuit consolidated the case with other cases involving similar facts.[408] In the case of *Gonzalez*, the court affirmed the lower court's decision to dismiss the case based on § 230.[409] The

[398] Lavi, *Do Platforms Kill? supra* note 44, at 522; *Gonzalez*, 282 F. Supp. 3d at 1153–54.

[399] Lavi, *Do Platforms Kill? supra* note 44, at 522.

[400] *Id.*; *Gonzalez*, 282 F. Supp. 3d at 1155.

[401] *Gonzalez*, 282 F. Supp. 3d at 1167–69; Plaintiff's Opposition to Google's Motion to Dismiss at 26–27, *Gonzalez*, 282 F. Supp. 3d 1150 (No. 4:16-cv-03282).

[402] *Gonzalez*, 282 F. Supp. 3d at 1169.

[403] *Id.* at 1171.

[404] *Id.* at 1169.

[405] *Id.*

[406] *Id.* at 1168–71.

[407] *Id.* at 1168.

[408] Gonzalez v. Google LLC, 2 F.4th 871, 871 (9th Cir. 2021), *vacated*, 598 U.S. 617 (2023), *remanded*, 71 F.4th 1200 (9th Cir. 2023).

[409] *Id.* at 895–96, 913.

majority found that "Google did not 'materially contribute' to its own recommendations."[410] By adopting a narrow reading of "information content provider," the court extended § 230 immunity to internet service providers that create or develop information through algorithmic processes.[411] Judge Berzon, though concurring with the majority, opined that were he not bound by Ninth Circuit precedent, he would agree with Judge Katzmann in *Force* and called for a more limited reading of the scope of § 230 immunity: "the term 'publisher' under § 230 reaches only traditional activities of publication and distribution—such as deciding whether to publish, withdraw, or alter content—and does not include activities that promote or recommend content or connect content users to each other."[412]

The U.S. Supreme Court granted certiorari in *Gonzalez* on whether immunity should apply to targeted recommendations of information provided by other information content providers.[413] The Court sent the case back to the Ninth Circuit, citing its decision in *Taamneh*.[414] Because the *Taamneh* case was dismissed on the basis of substantive law, the *Gonzalez* court did not discuss § 230.[415]

[410] Vincent Dumas, Comment, *Enigma Machines: Deep Learning Algorithms as Information Content Providers Under Section 230 of the Communications Decency Act*, 2022 WIS. L. REV. 1581, 1598 (2022); *see also* Cyphert & Martin, *supra* note 43, at 175–76 (discussing the *Gonzalez* case).

[411] Tomer Kenneth & Ira Rubinstein, Gonzalez v. Google*: The Case for Protecting "Targeted Recommendations"*, 72 DUKE L.J. ONLINE 176, 191 (2023) ("On this view, Section 230 immunity extends to using recommendation algorithms to match content and users, regardless of the outcomes."). Notably the sole exception for the immunity was the plaintiffs' allegations that Google approve ISIS videos for advertisement and shared proceeds with ISIS through YouTube's revenue sharing system. *Gonzalez*, 2 F.4th at 907. The Ninth Circuit held that these potential claims were not barred by § 230, but that plaintiffs failed to state a viable claim. *Id.*

[412] *Gonzalez*, 2 F.4th at 913.

[413] Gonzalez v. Google LLC, 598 U.S. 617, 621–22 (2023), *remanded to* 71 F.4th 1200 (9th Cir. 2023).

[414] *Id.* (citing Twitter, Inc. v. Taamneh, 598 U.S. 471 (2023)); Hyemin Han, *Supreme Court Rules in Favor of Twitter in* Taamneh*, Remands* Gonzalez, LAWFARE (May, 18, 2023, 11:13 AM), https://www.lawfaremedia.org/article/supreme-court-rules-in-favor-of-twit
ter-in-taamneh-remands-gonzalez [https://perma.cc/J2U2-R2VH]; Eric Goldman, *The Internet Survives SCOTUS Review (This Time)*—Twitter v. Taamneh *and* Gonzalez v. Google, TECH. & MKTG. L. BLOG (May 18, 2023) [hereinafter Goldman, *The Internet Survives SCOTUS Review*], https://blog.ericgoldman.org/archives/2023/05/the-internet-survives-scotus-review-this-time-twitter-v-taamneh-and-gonzalez-v-google.htm [https://perma.cc/W36J-GEV8].

[415] *Gonzalez*, 598 U.S. at 621.

Taamneh focused on whether Twitter supported terrorism and did not directly address § 230.[416] The family of Nawras Alassaf, who was killed in an ISIS attack in an Istanbul nightclub in 2017, sued Twitter under the Justice Against Sponsors of Terrorism Act ("JASTA").[417] The plaintiffs argued that Twitter knew the platform aided ISIS and could have taken "more aggressive action to combat pro-ISIS content posted on their sites."[418] Specifically, the plaintiffs argued that Twitter's recommendation algorithms extended "beyond passive aid and constitute[] active, substantial assistance."[419] The Ninth Circuit ruled in favor of the plaintiffs, accepting that the plaintiffs had plausibly alleged that the defendants aided and abetted ISIS within the meaning of the applicable statute.[420]

The U.S. Supreme Court then granted certiorari.[421] Commentators wondered whether the Court would impose liability on Twitter because only Justice Clarence Thomas had previously expressed a view on § 230, and he had criticized sweeping immunity for social media.[422]

[416] *Taamneh*, 598 U.S. at 506–07.

[417] *Id.* at 478–80, 483–84.

[418] *See* Alissa Donovan, *The Uncertain Fate of Section 230*, CARDOZO ARTS & ENT. L.J. BLOG (Apr. 3, 2023), https://larc.cardozo.yu.edu/cgi/viewcontent.cgi?article=1350&context=aelj-blog [https://perma.cc/9EEQ-SBE5].

[419] *Taamneh*, 598 U.S. at 499.

[420] *Id.* at 482.

[421] *Id.*

[422] John Fritze, *As Supreme Court Takes up Google Case, Only Clarence Thomas Has Made His Thoughts Clear*, USA TODAY (Jan. 31, 2023, 4:28 pm), https://www.usatoday.com/story/news/politics/2023/01/31/google-section-230-supreme-court-clarence-thomas/11149938002 [https://perma.cc/78XH-L7KB]; *see, e.g.*, Biden v. Knight First Amend. Inst., 141 S. Ct. 1220, 1224–27 (2021) (Thomas, J., concurring) (mem.) (comparing digital platforms to common carriers and suggesting that Congress might be able to pass "laws that restrict the platform's right to exclude"); Malwarebytes, Inc. v. Enigma Software Grp., LLC, 141 S. Ct. 13, 18 (2020) (mem.), *denying cert. to* 946 F.3d 1040 (9th Cir. 2019) (explaining the denial of certiorari where Justice Thomas noted that "[e]xtending § 230 immunity beyond the natural reading of the text can have serious consequences"); Cyphert &. Martin, *supra* note 43, at 176 ("Justice Thomas appeared on several occasions to be inviting lower court judges to reconsider the broad and sweeping nature of the holding in cases" granting immunity to platforms."); Balkin, *Free Speech*, *supra* note 100, at 1231 (noting that "Justice Clarence Thomas has suggested that large social media companies might be treated as common carriers or as public accommodations"); Goldman, *The Internet Survives SCOTUS Review*, *supra* note 414 ("It's frankly a little shocking to see Justice Thomas come out swinging in favor of algorithms, but here we are.").

However, with Justice Thomas delivering the opinion, the Court unanimously ruled in favor of Twitter.[423]

The case was decided based on anti-terrorism statutes, narrowly interpreting the statutes' requirements and using factors taken from before the digital age.[424] The Court used these factors not as bright-line rules but as guideposts to help determine that Twitter had not engaged in 'conscious, voluntary, and culpable participation' in the ISIS attacks.[425] In doing so, the Court found that the plaintiff's allegations satisfied some of the factors by alleging that ISIS committed a wrong and that the defendants knew they had a role in ISIS' activities but had failed to allege that the defendants "gave such knowing and substantial assistance to ISIS that they culpably participated in the Reina attack."[426] Although the Court concluded that Twitter knew it was playing a role in ISIS's enterprise, the plaintiffs failed to prove scienter and substantial assistance.[427] The Court explained that:

> The mere creation of those platforms, however, is not culpable . . . But the same could be said of cell phones, email, or the internet generally. . . . As presented here, the algorithms appear agnostic as to the nature of the content, matching any content (including ISIS' content) with any user who is more likely to view that content. The fact that these algorithms matched some ISIS content with some users thus does not convert defendants' passive assistance into active abetting.[428]

[423] *Taamneh*, 598 U.S. at 477, 507.

[424] *Id.* at 483–94, 505–06 (describing the factors as "(1) 'the nature of the act assisted,' (2) the 'amount of assistance' provided, (3) whether the defendant was 'present at the time' of the principal tort, (4) the defendant's 'relation to the tortious actor,' (5) the 'defendant's state of mind,' and (6) the 'duration of the assistance' given" (quoting Halberstam v. Welch, 705 F.2d 472, 488 (D.C. Cir. 1983))). Factor three, "whether the defendant was 'present at the time' of the principal tort," was not outlined with the internet age in mind. *Id.* (quoting *Halberstam*, 705 F.2d at 488).

[425] *Id.* at 493.

[426] *Id.* at 497.

[427] *Id.* at 497–98. For criticism of this approach, see Margot E. Kaminski & Meg Leta Jones, *Constructing AI Speech*, 133 YALE L.J.F. 1212, 1238 (2024) ("If courts hew to these kinds of strict intent requirements, then AI 'speakers' would get off the hook where human speakers would not. This would perversely incentivize more otherwise unlawful speech by AI systems—protecting more speech generated by AI than speech by actual humans." (internal citations omitted)).

[428] *Taamneh*, 598 U.S. at 499. Notably that Court was inaccurate on this point. As Professor Eric Goldman explains, "[n]ot only do algorithms routinely handle different types of content differently (i.e., photos, text, and videos are all processed differently),

The Court adopted a narrow interpretation of JASTA.[429] Accordingly, if the intermediary is not intentionally optimized to help terrorists, the plaintiffs cannot prove aiding and abetting terror, as they did not show sufficient assistance and scienter.[430]

> [The] 'defendants' platforms are global in scale and allow hundreds of millions (or billions) of people to upload vast quantities of information on a daily basis. Yet, there are no allegations that defendants treated ISIS any differently from anyone else. Rather, defendants' relationship with ISIS and its supporters appears to have been the same as their relationship with their billion-plus other users: arm's length, passive, and largely indifferent . . . . [P]laintiffs point to no act of encouraging, soliciting, or advising the commission of the Reina attack that would normally support an aiding and-abetting claim.[431]

The Court read the term knowingly "aiding and abetting" narrowly to require substantial assistance by the defendant in carrying out the terrorist attack.[432] The plaintiffs failed to identify a duty to terminate the service for bad actors that use it for illicit ends.[433] The Court acknowledged that there may be situations where such a duty exists but left the question open as to what situations might give rise to such a duty.[434] The Court noted that, even if there were a duty of care, the defendant's distant inaction would not rise to the level of aiding and abetting[435] because the petitioners failed to demonstrate that Twitter intentionally provided any substantial aid to the terror attacks or otherwise consciously participated in terror attacks, concluding that "once the platform and sorting-tool algorithms were up and running, defendants at most allegedly stood back and watched; they are not alleged to have taken any further action with respect to ISIS."[436]

but algorithms account for the content's substance . . . . [S]ocial media services constantly iterate their algorithms in response to a variety of pressures." Goldman, *The Internet Survives SCOTUS Review*, *supra* note 414.

[429] *Taamneh*, 598 U.S. at 498–99, 505–06.

[430] *Id.* at 505–06.

[431] *Id.* at 500; *see also* Goldman, *The Internet Survives SCOTUS Review*, *supra* note 414.

[432] *Taamneh*, 598 U.S. at 493.

[433] *Id.* at 501 (suggesting that a platform can only be culpable for a failure to act if they are violating a specific legal duty).

[434] *Id.*

[435] "[I]naction cannot create liability as an aider and abettor." *Id.* at 491 (quoting Zoelsch v. Arthur Andersen & Co., 824 F.2d 27, 36 (D.C. Cir. 1987)). Notably, the Court did not address the fact that targeting is beyond a failure to remove the bad actor but rather the defendant's own active behavior. *See id.*

[436] *Id.* at 499.

Subsequently, the Court turned to the causal connection. The Court held that the plaintiff's claim was overbroad and that accepting such a claim would mean any victim of an ISIS terrorist attack could bring the same claim.[437] Justice Thomas, however, addressed hypothetical situations in which providing a service could be considered aiding and abetting,[438] for example, providing dangerous wares to terrorist groups[439] or consciously or selectively promoting content provided by a particular terrorist group.[440] In such cases, a showing of more "direct, active, and substantial" aid would allow a plaintiff to establish liability.[441] While some bad actors abused the platform, that was insufficient to establish a claim that the defendants knowingly gave substantial assistance to terrorists.[442] Merely knowing that the wrongdoers were using the service and failing to stop them was not enough. The plaintiffs failed to allege a definable causal connection between the defendants' assistance and the *specific* Reina attack, and this fact "drastically increases their burden to show that defendants somehow consciously and culpably assisted the attack."[443] Without specific causal connection to the Reina attack, the plaintiffs' claims fell short of establishing a case for aiding and abetting the attack.[444]

Regarding the YouTube revenue sharing system, the Court ruled that the plaintiffs "allege[d] nothing about the amount of money that Google supposedly shared with ISIS, the number of accounts approved for revenue sharing, or the content of the videos that were approved."[445] Thus, the plaintiffs failed to prove that sharing such revenues constituted substantial assistance to terror.[446]

The Supreme Court reversed the Ninth Circuit's ruling on the basis of § 2333(d)(2),[447] noting the plaintiffs failed to prove the elements of

[437] *Id.* at 502.

[438] *Id.*

[439] *Id.* ("There may be, for example, situations where the provider of routine services does so in an unusual way or provides such dangerous wares that selling those goods to a terrorist group could constitute aiding and abetting a foreseeable terror attack.").

[440] *Id.*

[441] *Id.*

[442] *Id.* at 503.

[443] *Id.*

[444] *Id.* at 505.

[445] *Id.*

[446] *Id.*

[447] *Id.* at 506–07.

knowingly providing substantial assistance to terror and lack of sufficient causal connection.[448]

Taamneh did not address § 230, nor did it entirely close the door on the idea of liability for targeting.[449] The ruling focused on anti-terrorism statutes and their elements; the result might have been different under other torts.[450] Indeed, Justice Thomas narrowly interpreted intermediary liability for sorting-tool recommendation algorithms, addressing liability only from the stage when the platform is up and running.[451] However, a stage of system design and programming was neglected in *Taamneh*, and this will be addressed in the following Section.[452]

e. Algorithmic targeting after SCOTUS ruling

In *Twitter, Inc. v. Taamneh*, the U.S. Supreme Court focused on the substantive liability of platforms for their algorithmic recommendations in the case of material support for terror and on the elements of antiterrorism statutes, such as scienter.[453] Thus, even in the wake of this decision, the scope of § 230 and the boundaries of liability for algorithmic recommendations remain unelaborated.[454] The ambiguity is even greater outside the context of antiterrorism statutes.[455]

Post *Taamneh*, courts continue to tackle § 230 and consider imposing liability for algorithmic targeting. Recently, in *Vargas v. Facebook, Inc.*,[456]

[448] *Id.* at 505–06.

[449] F. Paul Pittman, Hope Anderson & John Oltean, *Supreme Court Declines to Reconsider Foundational Principles of Internet Platform Liability*, WHITE & CASE (June 15, 2023), https://www.whitecase.com/insight-alert/supreme-court-declines-reconsider-foundational-principles-internet-platform-liability [https://perma.cc/55AM-YN3P] ("[T]he underlying allegations in *Gonzalez* are 'materially identical' to those in *Taamneh*, the Ninth Circuit may dismiss the case for failing to state a claim under Section 2333(d)(2) of the Anti-Terrorism Act—leaving the Section 230 question unresolved for now.").

[450] *E.g.*, *Taamneh*, 598 U.S. at 477–79 (explaining that the applicable statute in this case is § 2333).

[451] *Id.* at 499 ("[O]nce the platform and sorting-tool algorithms were up and running, defendants at most allegedly stood back and watched; they are not alleged to have taken any further action with respect to ISIS.").

[452] *See infra* Part III.

[453] *Taamneh*, 598 U.S. at 500.

[454] Pittman et al., *supra* note 449.

[455] *See id.* (discussing that the Court is unlikely to expand the nature of internet platform liability).

[456] No. 21-16499, 2023 WL 4145434 (9th Cir. June 23, 2023).

the Ninth Circuit returned to *Roommates.com* in its analysis of whether Facebook is a developer.[457] The case focused on discrimination and targeting that hinders opportunities.[458] The plaintiff, a disabled Hispanic female, a single parent from New York City, and a frequent Facebook user, posted photos of herself and her children.[459] Thus, the plaintiff alleged the platform could deduce from her photos that she was "a single parent, disabled female of Hispanic descent."[460] She sought housing in Manhattan, yet Facebook searches did not yield advertisements for housing there.[461] Subsequently, she searched for housing via a Caucasian friend's Facebook Marketplace, using identical search criteria;[462] the friend received more housing advertisements.[463] The plaintiff therefore alleged that Facebook's "targeting methods provide tools to exclude women of color, single parents, persons with disabilities and other protected attributes."[464] Similar to *Roommates.com*, the plaintiff alleged that such targeting tools narrowed her opportunities to view advertisements for housing.[465] The Ninth Circuit did not apply § 230 since the plaintiffs' claims focused on Facebook's conduct as a co-developer of content and not as a publisher of information[466]: "Facebook applies its own algorithms to its vast store of data to determine which categories apply to a particular user."[467] By using information collected from users, Facebook inferred other protected attributes.

> Facebook knew that Plaintiff Vargas fell within the categories of single parent, disabled, female, and of Hispanic descent. . . . For other attributes, Facebook applies its own algorithms to its vast store of data to determine which categories apply to a particular user. . . . Facebook was "much more than a passive transmitter of

[457] *Id.* at *3–4 (comparing the website's actions in a previous Ninth Circuit case, *Fair Housing Council v. Roommates.com*, to those of Facebook in the present Ninth Circuit case).

[458] For further information on targeting and discrimination, see Kim, *Manipulating Opportunity*, *supra* note 66, at 883–85.

[459] *Vargas*, 2023 WL 4145434, at *1.

[460] *Id.*

[461] *Id.*

[462] *Id.*

[463] *Id.*

[464] *Id.*

[465] *Id.* at *2.

[466] *Id.*

[467] *Id.*

> information provided by others; it [was] the developer, at least in part, of that information."[468]

Facebook argued that the tools were "neutral" because they were offered to all advertisers, not just housing advertisers.[469] However, the Ninth Circuit ruled that the tools were discriminatory as they allowed targeting advertisements based on protected demographic criteria, and a discriminatory tool does not become neutral just because it is offered to others.[470]

Although *Vargas* focuses on § 230,[471] the very understanding that the social media platform can be considered a co-developer in the case of targeting is a step towards imposing liability.[472] Recently, in a case regarding the death of children who performed the TikTok "Blackout Challenge,"[473] which encourages children to choke themselves until passing out, parents sued social media platforms[474] claiming that the platforms were aware of the dangerous challenges yet promoted them to vulnerable children.[475] The Superior Court of the State of California ruled that laws protecting free speech and § 230 do not prevent negligence claims from proceeding.[476] Social media companies could, therefore, be held liable for the allegations because they are "based on the fact that the design features of the platforms . . . not the specific content viewed by plaintiffs."[477] Indeed, understanding this case as one of direct liability for negligence paves the way to imposing liability.[478]

2. *Legislative efforts to narrow the application of § 230 CDA with respect to*

[468] *Id.* at *2–3.

[469] *Id.* at *3.

[470] *Id.*

[471] *Id.* at *2.

[472] *See supra* note 431 and accompanying text.

[473] Nate Raymond, *Mother Whose Child Died in TikTok Challenge Urges US Court to Revive Lawsuit*, REUTERS (Jan. 17, 2024, 2:38 PM), https://www.reuters.com/legal/mother-whose-child-died-tiktok-challenge-urges-us-court-revive-lawsuit-2024-01-17 [https://perma.cc/MR97-HHTY].

[474] *Id.*

[475] *Id.*

[476] *TikTok Lawsuit for Teenage Harm*, SOCIAL MEDIA VICTIMS L. CTR., https://socialmediavictims.org/tiktok-lawsuit [https://perma.cc/RKK8-F4RB] (last updated Apr. 24, 2024).

[477] Rosenblatt, *supra* note 42 (quoting Judge Carolyn B. Kuhl).

[478] *Id.* ("A state judge . . . said she'll allow the lawsuits to advance based on a claim that the companies were negligent—or knew that the design of their platforms would maximise minors' use and prove harmful.").

targeting

While courts are generally reluctant to narrow § 230 and hold intermediaries accountable for targeting, § 230 occupies Congress, as both Democrats and Republicans seem to agree reform is needed.[479] Specifically, lawmakers are crafting bills to narrow § 230's application, *inter alia*, to address algorithmic targeting.[480]

The Protecting Americans from Dangerous Algorithms Act[481] aims to amend § 230 and remove immunity for large tech platforms that distribute algorithmic recommendations of radical content that incites violence.[482] Specifically, the bill aims to exclude from immunity recommendations of materials released by terrorist groups or that promote conspiracies to interfere with civil rights while preserving immunity for other user-generated content.[483]

The Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms Act (the Safe Tech Act)[484] is another example of an effort to address targeting.[485] This bill excludes advertisements or other paid content and social media companies that "enabl[e] cyber-stalking, targeted harassment, and discrimination on their platforms" from § 230 immunity.[486]

[479] Cyphert & Martin, *supra* note 43, at 158.

[480] For a list of states with pending AI bills, see *Artificial Intelligence 2023 Legislation*, NAT'L CONF. OF STATE LEGISLATURES, https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation [https://perma.cc/68QM-V4TX] (last updated Jan. 12, 2024).

[481] H.R. 2154, 117th Cong. (2021).

[482] Press Release, Congresswoman Anna G. Eshoo, Reps. Eshoo and Malinowski Introduce Bill to Hold Tech Platforms Liable for Algorithmic Promotion of Extremism (Oct. 20, 2020), https://eshoo.house.gov/media/press-releases/reps-eshoo-and-malinowski-introduce-bill-hold-tech-platforms-liable-algorithmic [https://perma.cc/FJ84-CE4D].

[483] Eugene Volokh, *§ 230 and the Protecting Americans Against Dangerous Algorithms Act*, VOLOKH CONSPIRACY (Dec. 1, 2021, 7:11 PM), https://reason.com/volokh/2021/12/01/%C2%A7-230-and-the-protecting-americans-against-dangerous-algorithms-act [https://perma.cc/AX69-VWCJ].

[484] H.R. 3421, 117th Cong. (2021); Cyphert & Martin, *supra* note 43, at 177.

[485] H.R. 3421.

[486] Cyphert & Martin, *supra* note 43, at 177 (quoting Press Release, Mark R. Warner, U.S. Sen. From Va., Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230 (Feb. 5, 2021), https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230 [https://perma.cc/SV8H-WSWB]). The bill was introduced on February 28, 2023, in both the House and the Senate. Wayne Rash, *Lawmakers Introduce*

Similarly, the Justice Against Malicious Algorithms Act ("JAMA")[487] exempts certain technological uses from § 230 liability.[488] The bill seeks to limit the protection available to social media platforms when they recklessly recommend harmful content through algorithmic amplification.[489] JAMA also eliminates § 230 immunity for personalized recommendations of information that materially contribute to physical or severe emotional injury.[490]

Recently, Senators Josh Hawley and Richard Blumenthal introduced the No Section 230 Immunity for AI Act.[491] This bill would exempt AI from § 230 protection;[492] it clarifies that § 230 immunity would not apply to claims based on generative AI, aiming to protect users from harmful content produced by the latest advancements in AI technology.[493] This bill leaves targeting by AI algorithms completely outside the scope of immunity.[494] Furthermore, this bill could help encourage companies to heavily invest in mitigation strategies because otherwise, they would be exposed to liability.[495]

In summary, § 230 "ensures that no duty of care ever emerges in a vast range of online scenarios and eliminates the incentives for the best-positioned party to develop responses to avoid foreseeable risks of harm."[496] However, it could be narrowed by interpretation, at least in

Bipartisan Reforms Making Big Tech Liable as Supreme Court Considers Section 230, FORBES (Feb. 28, 2023, 5:01 PM), https://www.forbes.com/sites/waynerash/2023/02/28/section-230-reforms-introduced-into-us-senate-and-house-of-representatives [https://perma.cc/Z76L-FGDZ].

[487] H.R. 5596, 117th Cong. (2021).

[488] *Id.* § 2(a) (creating exemptions for small interactive computer services, user-specified searches, and providers of infrastructure services); Lavi, *Manipulating*, *supra* note 53, at 325.

[489] Lavi, *Manipulating*, *supra* note 53, at 325.

[490] H.R. 5596 § 2(a).

[491] S. 1993, 118th Cong. (2023); Katie Paul, *Bipartisan U.S. Bill Would End Section 230 Immunity for Generative AI*, REUTERS (June 14, 2023, 1:43 PM), https://www.reuters.com/technology/bipartisan-us-bill-would-end-section-230-immunity-generative-ai-2023-06-14 [https://perma.cc/2PEC-Q8KT].

[492] S. 1993 § 1 (excluding "claims and charges related to generative artificial intelligence" from § 230 immunity).

[493] Paul, *supra* note 491.

[494] S. 1993 § 1.

[495] *See* Henderson et al., *supra* note 41, at 619, 626 (analyzing how harm mitigation can decrease artificial intelligence's liability if § 230 does not apply).

[496] Citron, *How to Fix Section* 230, *supra* note 84, at 728 (quoting Hearing on Holding Big Tech Accountable: Targeted Reforms to Tech's Legal Immunity Before Subcomm. on Commc'ns & Tech. of the H. Comm. on Energy and Com., 117th Cong. 5 (2021)).

the context of targeting. Moreover, many lawmakers agree that § 230 should be narrowed.[497]

Beyond § 230, many believe that in some cases of targeting, the law should impose substantive liability. The following Part outlines a framework for imposing liability for negligent design that would address targeting.

III. TAKING INFLUENCE SERIOUSLY: OUTLINING A FRAMEWORK TO IMPOSE A DUTY OF CARE FOR NEGLIGENT ALGORITHMIC DESIGN THAT PUTS CHILDREN'S SAFETY AT RISK

This Part makes the case for imposing liability for algorithmic recommendations and targeting of children as a special context that establishes a duty of care. It outlines a framework for imposing liability and will demonstrate that the idea of imposing liability for targeting is particularly justified in the algorithmic society. This idea is not novel, and even legislators recognize the need to regulate targeting children. The following Sections will expand on legal efforts to address targeting and negligent design.

A. First Steps in Regulation: Recognizing Algorithmic Targeting and Platform Design as a Special Context

Long ago, regulators realized that children need stronger protection than the general population. In 1998, Congress enacted the Children's Online Privacy Protection Act (COPPA)[498] to protect children under thirteen from unfair or deceptive acts or practices regarding personal information.[499] COPPA applies to online service providers that knowingly collect personal data from children under thirteen.[500] *Inter alia*, the law requires websites to include a notice indicating what information is collected, how it is used, and the website's information disclosure practices.[501] Sites also must obtain verifiable parental

[497] Cyphert & Martin, *supra* note 43, at 177–78 ("[T]here is bipartisan support for amending Section 230, though little agreement on what that amendment would look like.").

[498] 15 U.S.C. §§ 6501–06.

[499] § 6502(a) (regulating the manner in which "operator[s] of a website or online service directed to children . . . collect personal information from a child"); *see* Eldar Haber, *The Internet of Children: Protecting Children's Privacy in a Hyper-Connected World*, 2020 U. ILL. L. REV. 1209, 1224–25 (2020).

[500] § 6502(a)(1).

[501] § 6502(b)(1)(A)(i); 16 C.F.R. § 312.4 (2019); Haber, *supra* note 499, at 1224–25.

consent for collecting, using, or disclosing children's personal information;[502] adhere to data retention and deletion requirements;[503] and refrain from offering prizes to children for disclosing more information.[504] Finally, the law requires websites to contain reasonable procedures to protect children's personal information.[505]

The law recognizes the special context of children, but it lags with respect to regulating the targeting of children.[506] Nevertheless, legislators are working to narrow this gap and promote bills aiming to minimize the negative effects of algorithmic targeting. Other bills present a more far-reaching approach and seek to ban children from social media altogether[507] by requiring platforms to verify users' ages.[508]

Recently, federal legislators introduced the Children and Teens' Online Privacy Protection Act (COPPA 2.0)[509] to expand COPPA protections to children under sixteen.[510] The bill applies to websites "directed to children" and requires inclusion of a broader set of sites.[511]

[502] 16 C.F.R. § 312.5; Haber, *supra* note 499, at 1225.

[503] 16 C.F.R. § 312.10; Haber, *supra* note 499, at 1224–25.

[504] 16 C.F.R. § 312.7; Haber, *supra* note 499, at 1224–25.

[505] 16 C.F.R. §§ 312.3(e), 312.6, 312.8; Haber, *supra* note 499, at 1224–25.

[506] *See generally* Haber, *supra* note 499 (explaining how protective regulatory frameworks that target online privacy threats apply to devices directed toward children; however, these regulations do not apply to devices that may inadvertently monitor children).

[507] SCOTT BABWAH BRENNEN & MATT PERAULT, KEEPING KIDS SAFE ONLINE: HOW SHOULD POLICYMAKERS APPROACH AGE VERIFICATION? 13 (2023); Social Media Child Protection Act, H.R. 821 § 2, 118th Cong. (2023) (requiring social media companies to prohibit children from using their platforms and requiring platforms to assure that users provide a government-issued ID to access the platform or use some other "reasonable method of verification," taking into account existing technology); Making Age-Verification Technology Uniform, Robust, and Effective Act, S. 419 § 2, 118th Cong. (2023) (requiring social media platforms to verify that account holders are older than sixteen); Protecting Kids on Social Media Act, S. 1291 §§ 4, 5, 118th Cong. (2023) (barring social media platforms from allowing users under the age of thirteen and requiring parental consent for users between thirteen and eighteen). Intermediaries would be banned from using algorithms to recommend content to young users. *Id.*

[508] BRENNEN & PERAULT, *supra* note 507, at 14; H.R. 821 § 2 (requiring social media platforms to verify users' age through government-issued identification or other "reasonable method of verification[s]"); S. 419 § 2 (specifying that social media platform verification requires potential users to upload a copy of government-issued identification); S. 1291 § 3 (instructing social media platforms to implement age verification processes for users).

[509] S. 1418, 118th Cong. (2023).

[510] *Id.* § 2(a)(17), (b)(1); BRENNEN & PERAULT, *supra* note 507, at 14.

[511] S. 1418 § 2(b); BRENNEN & PERAULT, *supra* note 507, at 13.

In addition, COPPA 2.0 would no longer require actual knowledge that children use the service.[512]

The Kids Online Safety Act (KOSA)[513] is another federal bill that establishes a "duty of care" for intermediaries to protect children from mental harm, sexual exploitation, and illegal substances.[514] The bill further requires companies to undergo independent external audits, imposes transparency requirements to allow researcher access to platform data assets, and creates substantial youth and parental controls to create a safer digital environment.[515]

Another bipartisan bill, introduced by California State Assembly members, proposed an amendment to the California Civil Code allowing parents and the state Attorney General to sue social media platforms if their child becomes addicted to the platform.[516] The bill exempted social media platforms from liability provided that it identified and eliminated its addictive features within thirty days after an audit.[517] However, the California State Senate Appropriations Committee halted this bill as it ignored the positive effects social media can have on teenagers.[518] In addition to ignoring social media's positive impact, the bill included loopholes for smaller websites.[519]

On September 25, 2022, California passed a new law—the California Age-Appropriate Design Code Act (CAADCA),[520] which will take effect

[512] S. 1418 § 2(b) (expanding the statute to apply not only to websites directed toward children but to websites with "actual knowledge or knowledge fairly implied" that children utilize their service); BRENNEN & PERAULT, *supra* note 507, at 13.

[513] S. 3663, 117th Cong. (2022).

[514] S. 3663 § 3; Amy Novotney, *Kids Online Safety Act: APA Leads More than 200 Advocates in Urging Senate to Pass Bill*, AM. PSYCH. ASS'N (July 18, 2023), https://www.apa.org/news/apa/2023/kids-online-safety-act-senate-letter [https://perma.cc/GV2Y-TGTE]; BRENNEN & PERAULT, *supra* note 507, at 13.

[515] S. 3663 §§ 4(a), 6, 7(b).

[516] Social Media Platform Duty to Children Act, Assemb. B. 2408, 2022 Leg., Reg. Sess. (Ca. 2022); Peter Suciu, *Social Media Liability Bill Passed California State Committee—Are Teens Actually 'Addicted' to Social Media?*, FORBES (May 6, 2022, 1:49 PM), https://www.forbes.com/sites/petersuciu/2022/05/06/social-media-liability-bill-passed-california-state-committeeare-teens-actually-addicted-to-social-media [https://perma.cc/PTF4-QEC8]

[517] Assemb. B. 2408; Suciu, *supra* note 516.

[518] *CA Senate Halts Social Media Addiction Bill*, CHAMBER OF PROGRESS (Aug. 11, 2022), https://progresschamber.org/ca-senate-halts-social-media-addiction-bill [https://perma.cc/F2N4-PC24].

[519] *Id.*

[520] Assemb. B. 2273, 2022 Leg., Reg. Sess. (2024 Ca.).

on July 1, 2024.[521] The Act will regulate the collection, storage, processing, and transfer of children's data[522] and will expand COPPA and California's Parent Accountability and Child Protection Act.[523] Whereas COPPA applies only to data of children under thirteen, CAADCA will apply to minors up to the age of eighteen.[524] Thus, it recognizes that children are susceptible to influence even beyond thirteen[525] and will require websites to offer a high level of privacy by design and by default to children under eighteen.[526]

Under CAADCA, businesses that develop and provide online services, products, or features that children will likely access will be required to prepare a "Data Protection Impact Assessment"[527] detailing how the feature's design could expose children to "potentially harmful" materials.[528] It will require social media companies to design their platforms and develop their codes with the best interests of child users in mind.[529] Section 1 (8) of the Act explicitly refers to child profiling and stipulates that "online services, products, or features that are likely to be accessed by children should offer strong privacy protections by design and by default, including by disabling features that profile children using their previous behavior, browsing history, or assumptions of their similarity to other children, to offer detrimental material."[530] Another important provision requires businesses to consider the best interests of children from the design stage.[531]

[521] *Id.*

[522] *Id.*

[523] Parent's Accountability and Child Protection Act, Assemb. B. 2511, 2018 Leg., Reg. Sess. (2020 Ca.). This Act prohibits a business from soliciting or knowingly permitting minors to consent on behalf of a parent or legal guardian, to use of the minor's name, image or information on a social media website or application, or to do so without a parent or legal guardian's permission. *Id.*

[524] CAL. CIV. CODE § 1798.99.30(b)(1).

[525] *Id.* § 1(a)(1); *cf.* Gilad et al., *supra* note 63, at 1297.

[526] Assemb. B. 2273 §§ 1, 2 (requiring websites to ensure individuals under eighteen have default privacy settings that confer maximum data privacy through completion of a Data Protection Impact Assessment that evaluates how the platform uses a child's personal information).

[527] Assemb. B. 2273 § 2; CAL. CIV. CODE §§ 1798.99.30(a)(2), 1798.99.33 (West 2023).

[528] Assemb. B. 2273 § 2.

[529] *Id.* (directing online providers to develop platforms in a manner that protects and upholds children's heightened privacy needs).

[530] *Id.*

[531] *Id.* § 2; CAL. CIV. CODE § 1798.99.29(a) (West 2021).

In order to implement the law in their policies, businesses will be required to estimate their users' age with "a reasonable level of certainty appropriate to the risks that arise from the data management practices of the business" or to "apply the privacy and data protections afforded to children to all consumers."[532] Thus, they will need to choose between age verification of all users or disabling part of their data collection features for all users. Such provisions would stop the deeply personalized targeting of children.[533]

Notably, CAADCA came under fire on December 14, 2022, when an online business trade association, NetChoice, sued the California Attorney General, challenging the constitutionality of the act.[534] NetChoice argued that CAADCA violates the First Amendment because it is vague and stipulates content-based regulation that compels companies to serve as "roving censors of speech."[535] Professor Eric Goldman, an Associate Dean at *Santa Clara University* and a renowned researcher of internet law and intellectual property, submitted an amicus brief in support of NetChoice, arguing that CAADCA "creates barriers for both minors and adults seeking to access websites or apps" and citing how those barriers impermissibly deter or even block "users from engaging in activities that are protected by the First Amendment."[536] Arguably, an age verification requirement would burden social media platforms and users by reducing users' platform consumption and contributions.[537]

Recently, a federal judge in San Jose blocked California from enforcing CAADCA by issuing a preliminary injunction because its

[532] Assemb. B. 2273 § 2.

[533] *Id.* (ensuring platforms cannot bypass the bill's heightened data privacy requirement by not performing adequate estimations of users' ages).

[534] NetChoice, LLC v. Bonta, No. 5:22-cv-08861-BLF, 2023 WL 6135551, at *1 (N.D. Cal. Sept. 18, 2023), *appeal docketed*, No. 23-02969 (9th Cir. Oct. 23, 2023); *see* Mengting Xu, *Lawsuit Challenges Constitutionality of California Age—Appropriate Design Code, California Lawyers Association*, CAL. LAW. ASS'N (Feb. 9, 2023), https://calawyers.org/privacy-law/lawsuit-challenges-constitutionality-of-california-age-appropriate-design-code [https://perma.cc/6PMA-2CRB].

[535] Xu, *supra* note 534; *NetChoice*, 2023 WL 6135551, at *1, *4.

[536] *See* Brief of Professor Eric Goldman as Amicus Curiae Supporting Plaintiff at 1, *NetChoice*, 2023 WL 6135551 [hereinafter Goldman Amicus Brief]; *Eric Goldman*, SANTA CLARA UNIV. SCH. OF L., https://law.scu.edu/faculty/profile/goldman-eric [https://perma.cc/R2Z7-JKTW].

[537] Goldman Amicus Brief, *supra* note 536, at 5. The Article will expand on these arguments *infra* Part IV.

speech restrictions likely violate the U.S. Constitution's First Amendment.[538] The final ruling remains to be seen.

Another bill addressing the special context of targeting children is Features that Harm Child Users: Civil Penalty.[539] This bill, authored by state Senator Nancy Skinner, defines child users as those below sixteen years old,[540] and it prohibits social media platforms from

> us[ing] a design, algorithm, or feature that the platform knows, or by the exercise of reasonable care should have known, causes a child user to do any of the following: (1) inflict harm on themselves or others; (2) develop an eating disorder; (3) experience addiction to the social media platform . . . .[541]

However, the bill outlines that social media platforms shall not be considered violators if they have both instituted and maintained a "program of at least quarterly audits of its designs, algorithms, and features that have the potential to cause violations of subdivision (a)," and "corrected within 60 days any design or algorithm discovered by the audit to present more than a *de minimis* risk of violating" subdivision a.[542]

This bill was also met with the criticism that it was overbroad and vague.[543] It was alleged that the broad definition of addiction could result in websites avoiding hosting child users from California, as every design feature could be perceived as addictive and expose them to liability.[544] Alternatively, websites might avoid promoting innovation or developing efficient algorithms that usually expose children to

[538] *See* Order Granting Motion for Preliminary Injunction at 37, 42, *NetChoice*, 2023 WL 6135551.

[539] S.B. 680, 2023 Leg., Reg. Sess. (2023 Cal.).

[540] *Id.*

[541] *Id.*

[542] *Id.* § 1714.48(b).

[543] *See* Austin Jenkins, *Calif. Lawmaker Aims to Ban Addictive Social Media Algorithms*, PLURIBUS NEWS (Feb. 13, 2023, 6:00 AM), https://pluribusnews.com/news-and-events/calif-lawmaker-aims-to-ban-addictive-social-media-algorithms [https://perma.cc/BK6U-CG7J] (quoting Carl Szabo, Vice President and General Counsel at NetChoice: "[The proposed bill] injects government and technology in between parents and their teenagers . . . . And, it does little to address the underlying issues raised by social media—responsible use of technology").

[544] *See* Press Release, *Sen. Skinner Introduces Landmark Bill to Protect Youth from Social Media Addiction*, SENATE DIST. 09 (Jan. 29, 2024), [hereinafter *Sen. Skinner Press Release*] https://sd09.senate.ca.gov/news/20240129-sen-skinner-introduces-landmark-bill-protect-youth-social-media-addiction [https://perma.cc/L66H-38HX] (reporting Senator Skinner's support for the bill barring social media platforms from sending any addictive social media content to children without a guardian's consent).

relevant content.[545] However, although this bill may be too vague and need refinement, addressing the special context of children and the harm targeting inflicts upon them is desirable.

Beyond specific bills at the federal and state levels, federal agencies have promulgated regulations to support children's online safety. The Federal Trade Commission (FTC) recently proposed changes that would bar social media from profiting from data collected from users under eighteen.[546] In addition, intermediaries would be prohibited from using information on children to target them with advertisements.[547] This regulation successfully considers the context of children in the algorithmic society.

B. A Duty of Care for Negligent Design: Children's Algorithmic Targeting as Negligent Design

As discussed throughout this Article, recommendation systems that target children create a special context of vulnerability. Such recommendations can give the target the impression that the system supports the content and validates its message or the source of the message.[548] Consequently, the message's entire context changes when children receive it directly without them actively searching for it.[549] These circumstances render the information far more influential and could lead to harm that stands apart from any that might have been inflicted by the content alone.[550] Exposing individuals repeatedly to specific types of content can eventually shape their behavior.[551]

Understanding targeting's cognitive influence on context, the flow of information, and children's behavior in particular, should lead policymakers to apply a just and efficient legal policy for intermediary

[545] *See id.* (discussing how the bill will target social media companies to prevent them from creating profitable algorithms that would promote addiction).

[546] Singer, *supra* note 218.

[547] *Id.*

[548] *See Sen. Skinner Press Release*, *supra* note 544 (describing how the California bill will prevent social media platforms from notifying children of the platforms' recommendations because of its psychological effects on youths).

[549] *See id.* (detailing Senator Skinner's concern over children receiving social media notifications at late hours or while in school).

[550] *See* MANHEIM & ATIK, *supra* note 270, at 10 (explaining that research indicates harms caused by social media algorithms may be different and separate from harms caused by recommended content).

[551] *See supra* Section III.A (recognizing the concern is magnified when recommendations are targeted towards children because of their unique susceptibility).

liability with respect to targeting children. This Article has demonstrated cases where creating a special context by design and other forms of influence can result in liability.[552] The previous Part recognized a duty of care for algorithmic design.[553] Indeed, companies are free to design their systems as they see fit, but they should be subject to basic requirements to keep children safe, and they should internalize the costs they impose on society through algorithmic recommendations and targeting children.[554] Mandating technology companies to uphold a duty of care would incentivize them to design safer products[555] and develop more technology and algorithms that would increase the efficiency and accuracy of the user's experience.[556] Imposing a duty of care is not a revolutionary idea; scholars have already discussed a duty of care for intermediaries in various contexts.[557]

1. Negligent design and liability

Beyond § 230 immunity, "courts have started to look far more carefully at the ways in which the designs of interactive computer services cause informational harm."[558] As described earlier, *Maynard* is a prominent case in which the court recognized a duty of care and allowed the imposition of liability for the design of a virtual speed

[552] *See supra* Section I.D.II.

[553] *See infra* Section III.B (outlining how a duty of care exists in case law, like *Maynard*, and the Kids Online Safety Act, and thus could apply to algorithmic design).

[554] WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 126 (2018).

[555] *See* Citron, *How to Fix Section 230*, *supra* note 84, at 727 (explaining that there is no legal incentive for platforms to take down illicit materials or identify predators).

[556] Lavi, *Do Platforms Kill?*, *supra* note 44, at 554.

[557] *See id.* (proposing a "duty of care" to algorithmic targeting); Citron, *How to Fix Section 230*, *supra* note 84, at 761 (discussing a proposed bill in the United Kingdom that would require online platforms to have a "duty of care" to improve online safety); Kim, *Beyond Section 230 Liability*, *supra* note 50, at 363 ("Given the special relationship that Facebook has with its users, which it uses for its financial benefit, the company has a duty to exercise reasonable care with respect to risks that arise within that relationship.").

[558] Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 YALE L.J.F. 475, 497 n.125 (2021) (referring to *Gonzalez* and *Lemmon*). Professor Sylvain also noted that in *FTC v. LeadClick Media*, LLC, 838 F.3d 158 (2d Cir. 2016), "[t]he Federal Trade Commission (FTC) . . . successfully argued, for example, that an online company that promoted false stories about nutritional supplements across a network of affiliated websites was not a mere 'publisher' within the meaning of Section 230." *Id.*

filter.[559] Recall that Georgia's Supreme Court remanded the case for further proceedings, reasoning that "Snap could reasonably foresee the particular risk of harm from the Speed Filter at issue . . . ."[560] The court has also recognized a particular manufacturer's legal duty of care to ensure that its products are not designed defectively.[561] The court recognized that application providers owe duties with respect to the design of their products.[562] Such an approach can be applied to algorithmic targeting, subjecting it to liability for negligent design.[563]

2. *A proposed framework for regulating the targeting of susceptible children*

This Section proposes a comprehensive framework aimed at regulating algorithmic targeting, especially in the context of safeguarding children from potentially harmful material. First, the framework provides for a total prohibition on targeting of children and an opt-in system for adults. Second, the framework recognizes several age verification methods available to intermediaries while emphasizing the importance of minimizing data security risks. Lastly, the framework outlines potential legal remedies for parents and children should intermediaries fall short of the proposed framework by failing to adhere to the duty of care.

a. Abstaining from targeting children

Targeting children under eighteen should be a breach of the duty of care. The proposed framework refers to age eighteen, not thirteen, because children's brains are still not ripe to make competent decisions. Even though experts doubt whether teens' brains are

[559] *See* Maynard v. Snapchat, Inc., 313 Ga. 533 (Ga. Sup. Ct. Mar. 15, 2022); *see supra* notes 343–352 and accompanying text.

[560] *Maynard*, 313 Ga. at 534.

[561] *Id.* at 542.

[562] *Id.*

[563] Eugene Volokh, *Large Libel Models? Liability for AI Output*, 3 J. FREE SPEECH L. 489, 522 (2023) (explaining that a company is "potentially responsible for harms caused by the equipment it uses in the course of business, at least when it negligently fails to take reasonable steps to minimize the risks of those harms" and this negligence standard should also apply to an AI company when it produces AI software that it knows communicates false and defamatory statements). This includes its algorithmic recommendation systems. *Id.*

sufficiently developed at eighteen,[564] in most jurisdictions, eighteen is the age threshold regarding adult decision-making capability.[565]

Unlike proposals in literature to prohibit commercial targeting and data collection altogether,[566] the proposed framework focuses on targeting of children and an opt-in approach for targeting of adults.[567]

b. A duty to implement age verification systems for companies that continue the practice of targeting algorithmic recommendations

Companies that continue targeting other populations would be required to set up an age verification system to exclude children from targeting. Accordingly, users (children and adults) who do not verify their age would not receive algorithmic recommendations. Data regarding the user's age, or whether the user is a child, would only be available to the intermediary.

This framework would exclude individuals who have not verified their age from targeting entirely, and this exclusion would not be limited to recommendations of harmful content, thereby preventing ambiguity regarding the definition of "harmful . . . content."[568] Moreover, the framework avoids content-based regulation, which might be struck down as unconstitutional.[569]

[564] *See* Gilad et al., *supra* note 63, at 1297 ("[S]etting the bar at age [thirteen] contradicts the scientific consensus that the prefrontal cortex, a brain region that plays an important role in decision-making and impulse control, is not fully developed before age [twenty-five].").

[565] *Id.*

[566] CARISSA VELIZ, PRIVACY IS POWER: WHY AND HOW YOU SHOULD TAKE BACK CONTROL OF YOUR DATA 141 (2020).

[567] For more details on this opt-in approach, see *infra* Part IV.

[568] CAADA, CA. CIV. CODE § 1798.99.31 (2024), requires companies to create a report that addresses "whether the design of [their] online product, service, or feature could harm children, including by exposing children to harmful, or potentially harmful, content" and "whether algorithms used . . . could harm children." *Id.* This vague definition leads to unclarity and came under fire and for that reason, *NetChoice* challenged this legislation on First Amendment grounds and sought an injunction blocking its implementation while under court review. *See generally* Complaint for Declaratory and Injunctive Relief, NetChoice, LLC v. Bonta, No. 5:22-cv-08861, 2023 WL 6135551 (N.D. Cal. Dec. 14, 2022).

[569] *See, e.g.*, *NetChoice*, 2023 WL 6135551, at *23, *appeal docketed*, No. 23-02969 (9th Cir. Oct. 23, 2023) (holding that NetChoice could likely succeed on a First Amendment theory challenging the constitutionality of the content-based CAADA).

Age verification is not a novel concept. It has been adopted in several bills in various contexts, such as for pornographic sites.[570] There are also bills that ban children from social media unless they obtain parental consent.[571] To do so, platforms will need to establish acceptable methods of identification.[572] Specifically, there are bills regarding algorithmic targeting, such as the aforementioned CAADCA, which imposes requirements on internet sites directed at or "likely to be accessed by children."[573]

Notably, the age verification requirement has also been adopted outside the United States. Recently, the United Kingdom (UK) enacted the Online Safety Act.[574] Under the act, internet platforms with UK users will be required to prevent minors from accessing 'harmful' content, as defined by the UK Parliament.[575] In order to do so, platforms will have to verify or estimate the age of their users.[576] Thus, adopting an age verification requirement in the United States is not far-fetched, and as this Article explains, such a requirement can align with the First Amendment.[577]

[570] BRENNEN & PERAULT, *supra* note 507, at 16; *see, e.g.*, Louisiana Act No. 440, 2022 Regular Session; Mississippi SB 2346, 2023 Regular Session; Tim Cushing, *Louisiana Law Now Requires Age Verification at Any Site Containing More than One-Third Porn*, TECHDIRT (Jan. 3, 2023, 12:05 PM), https://www.techdirt.com/2023/01/03/louisiana-law-now-requires-age-verification-at-any-site-containing-more-than-one-third-porn [https://perma.cc/LLB5-E4E4].

[571] Dustin Jones, *Kids Under 13 Would Be Barred from Social Media Under Bipartisan Senate Bill*, NPR (Apr. 28, 2023, 5:00 AM), https://www.npr.org/2023/04/28/1172098173/social-media-kids-senate-bill [https://perma.cc/ZDR3-MCJL].

[572] BRENNEN & PERAULT, *supra* note 507, at 15.

[573] California Age-Appropriate Design Code Act, CA. CIV. CODE § 1798.99.31 (2024); *see also Online Safety Act 2023*, c. 50 (Eng.), https://www.legislation.gov.uk/ukpga/2023/50/enacted [https://perma.cc/G88L-WMGB]; BRENNEN & PERAULT, *supra* note 507, at 15 ("[W]hile the upshot of the bill is that most sites will need to verify the age of users, it does not specify how they should do so."). This Article will further expand on bills related to targeting of children, including age verification, *infra* Section III.B.

[574] Online Safety Act 2023, c. 50.

[575] *See id.* § 12(4) ("The duty set out in subsection (3)(a) requires a provider to use age verification or age estimation (or both) to prevent children of any age from encountering primary priority content that is harmful to children which the provider identifies on the service.").

[576] *Id.* § 230.

[577] *See infra* Section III.C.3.

There are various ways to verify age. Some platforms require users to submit identification while others adopt a "risk-based approach."[578] Social media platforms can rely on an age declaration with third-party verification. For example, Instagram is testing a system of asking three friends of a user claiming to be over eighteen to confirm the user is in fact over eighteen.[579] Most social media platforms ask users to input their birth date when registering a new account.[580] Another method is an identifier submission requirement where users must upload images of official documents that include their birth dates,[581] making it harder for children to bypass the verification.[582] However, this approach involves the undesirable collection of sensitive information that can be breached, raising data security concerns.[583]

Another method is verification by a third-party institution, like a credit card company.[584] In the European Union (EU), there are efforts to create a digital identity mechanism that would "allow users to submit documents to a central system to verify their age."[585] In the surveillance capitalism era, intermediaries can also verify their users' age without explicitly asking by using "AI-based inference systems to assess user content and behavior to identify users who might fall under a certain age."[586] There can, however, be problems of accuracy with this

[578] BRENNEN & PERAULT, *supra* note 507, at 14 (noting that state online safety bills typically take one of these two approaches when it comes to age verification); *see* DANIEL J. SOLOVE & WOODROW HARTZOG, BREACHED 71–75 (2022) (explaining the need to balance comfort, functionality, and information security).

[579] Ivan Mehta, *Instagram Tests New Age Verification Tools, Including Video Selfies*, TECHCRUNCH (June 23, 2022, 6:00 AM), https://techcrunch.com/2022/06/23/instagram-tests-new-age-verification-tools-for-18-and-over-accounts-including-video-selfies [https://perma.cc/BSP6-P78P].

[580] BRENNEN & PERAULT, *supra* note 507, at 3, 5.

[581] *See, e.g.*, *Access Age-Restricted Content & Features*, GOOGLE, https://support.google.com/accounts/answer/10071085?hl=en [https://perma.cc/FKA4-HMNJ]; *How Does Age Verification Work?*, TINDER, https://www.help.tinder.com/hc/en-us/articles/360040592771-How-does-age-verification-work [https://perma.cc/UCY3-NB7F]; *How Video Selfie Age Verification Works on Facebook Dating*, FACEBOOK, https://www.facebook.com/help/661251112277115 [https://perma.cc/PVL5-N7YC].

[582] BRENNEN & PERAULT, *supra* note 507, at 3.

[583] *Id.* at 3, 8.

[584] *Id.* at 4.

[585] *Id.*

[586] *Id.* at 4–5.

method.[587] Additionally, this requires AI to assess user content and behavior, which may be intrusive, thereby making it a less desirable option.[588]

This proposed framework prefers positive requirements for users to submit identification or verify their identity via third parties; a user who declines to do so would not receive algorithmic recommendations. Indeed, security measures such as encryption or anonymization of personal information would be needed to mitigate the risks.[589] Yet, despite the potential costs of such measures, the option to submit identification is preferable to vague legislation that requires "estimat[ing] the age of child users" at a "reasonable level of certainty" without defining reasonable levels of certainty.[590]

c. Remedy

If a duty of care to avoid targeting children is not realized, a tort law remedy should be possible. Children and their families who received recommendations of harmful content, acted upon them, and suffered severe bodily harm or death could file a suit against the intermediaries. The question of remedy and the compensation amounts would depend on the facts of the case.[591]

Indeed, imposing liability for negligent design is not without difficulty, especially when AI algorithms allegedly cause the harm.[592] Several objections to the proposed framework are anticipated. There are questions regarding imposing liability on learning algorithms, questions of the causal connection between the design and harm, and

[587] *Id.* at 3–4. *See generally* Tzvi Ganel, Carmel Sofer & Melvyn A. Goodale, *Biases in Human Perception of Facial Age Are Present and More Exaggerated in Current AI Technology*, SCI. REPS. (Dec. 29, 2022), https://www.nature.com/articles/s41598-022-27009-w [https://perma.cc/59Y4-UWHE] (finding that AI technology is less accurate than human performance in estimating age through facial recognition).

[588] *See* BRENNEN & PERAULT, *supra* note 507, at 5 (highlighting that "expansive collection and processing of minors' data is precisely what some regulators are trying to prevent").

[589] *See* Hadar Y. Jabotinsky & Michal Lavi, *Speak Out: Verifying and Unmasking Cryptocurrency User Identity*, 32 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 518, 586–89 (2022) [hereinafter Jabotsinky & Lavi, *Speak Out*] (discussing encryption and data anonymization).

[590] Assemb. B. 2273 § 2.

[591] *See infra* Section III.C.3.

[592] *See infra* Section III.C.1.

First Amendment objections.[593] The following Sections address these questions and difficulties.

C. Addressing Potential Objections to the Proposed Framework

The following Sections predict and examine the potential criticisms of the proposed framework for managing and regulating AI-driven targeting. The first part confronts the question of liability in the context of predictive and autonomous capabilities, arguing for a nuanced understanding of intermediary liability. The second part discusses the causal nexus between algorithmic targeting and the resulting harm, a complex, but critical element to establish a successful tort claim against an intermediary. The last part considers targeting in a First Amendment context and argues that the proposed framework promotes children's privacy and safety and does not infringe upon fundamental free speech rights.

1. How can the law impose liability for AI targeting by autonomous algorithms?

> There are humans behind every algorithmic prediction, much like the Wizard of Oz was a man operating a machine.[594]

The first obstacle to recognizing a duty of care is that recommendation systems target content automatically using autonomous AI algorithms. The algorithmic learning process and the reactions of learning algorithms can be unpredictable.[595] Should intermediaries have a duty of care regarding such targeting? This Article answers this question with a resounding "yes." First, the proposed framework prohibits the targeting of children and individuals who have not verified their age. Thus, the intermediary controls whether the recommendation system can take effect. Deciding whether to allow the algorithm to operate is completely under the intermediary's control and differs from instructing the algorithm on what to target.[596]

However, even under a content-based framework that forbids targeting children with harmful content, intermediaries should not

[593] *See infra* Section III.C.

[594] MATSUMI & SOLOVE, *supra* note 75, at 11.

[595] Lavi, *Do Platforms Kill?*, *supra* note 44, at 558.

[596] Jack M. Balkin, *2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1221 (2017) [hereinafter Balkin, *2016 Sidley Austin*].

hide behind autonomous algorithms and should not be exempt from implementing reasonable instructions to limit such recommendations. The rights and obligations of algorithms should be attributed to the intermediary, which can then bear responsibility for implementing negligent design.[597] After all, humans design, program, and connect these algorithmic systems to the database.[598] Humans develop and train the software to differentiate between more or less appropriate outputs.[599] In addition, humans decide how and when to use the algorithms and for what purpose.[600] In addition, humans can "limit the algorithmic learning process and teach algorithms to . . . mitigate harmful consequences of their usage."[601] Thus, although algorithms can self-learn, humans direct the purpose of technology.[602] The programmers can limit the parameters for self-learning *ex-ante* or block specific system results,[603] as well as design choices that ultimately affect the function and outcomes of these algorithmic systems.[604]

Limitations by design are applied in other technological contexts and can be transplanted into the context of algorithmic

[597] *Id.* at 1223.

[598] *Id.* at 1221–22; *see also* Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age*, PEW RSCH. CTR. 1, 11–13 (Feb. 8, 2017), https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age [https://perma.cc/G4JJ-D5FH] (highlighting the inherent biases in algorithmic systems partially due to its human creators).

[599] *See* Bambauer & Surdeanu, *supra* note 74, at 379 (explaining that ChatGPT is a "pleaser" because its responses are influenced by the "queries it receives").

[600] Balkin, *2016 Sidley Austin*, *supra* note 596, at 1223.

[601] Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 27.

[602] LOBEL, *supra* note 78, at 289 ("[D]esigning machines, directing their purposes, and defining the goals of algorithms is somethings we humans do."); *see also* Sharma Chinmayi, *AI's Hippocratic Oath*, WASH U. L. REV. (forthcoming) (manuscript at 44), https://ssrn.com/abstract=4759742 (arguing that engineers who design AI can address AI harms and that professionalization of engineers would separate the motivations of individual engineers from the corporate incentives fueling the ongoing AI race to the bottom, instilling a sense of social responsibility).

[603] *See* Alexei Dulub, *How to Avoid Bias in Machine Learning Algorithms*, FORBES (Apr. 27, 2023, 6:00 AM), https://www.forbes.com/sites/forbestechcouncil/2023/04/27/how-to-avoid-bias-in-machine-learning-algorithms/?sh=53621082352a [https://perma.cc/UG59-BT5T] (discussing PixelPlex's methods to control and avoid bias within machine learning algorithms).

[604] Andrew D. Selbst, Suresh Venkatasubramanian & I. Elizabeth Kumar, *Deconstructing Design Decisions: Why Courts Must Interrogate Machine Learning and Other Technologies*, 85 OHIO ST. L.J. (forthcoming 2024) (manuscript at 1), https://ssrn.com/abstract=4564304 [https://perma.cc/QX3X-362C].

recommendations.[605] For example, Apple's Siri has design limitations and "sidesteps medical, legal, or spiritual counsel."[606] Similarly, the Open AI developers have put safeguards on their device so that it will decline generating texts on certain controversial subjects.[607] The Cyberspace Administration of China specifically recognizes the power to restrict AI *ex-ante*;[608] China released a draft of Administrative Measures for Generative AI Services to "ensure that content created by generative AI is consistent with the 'social order and societal morals,' and 'does not endanger national security.'"[609] Although some may consider such limitations overbroad, this example demonstrates that limiting such systems is possible.[610] This is all the more true when an AI company has received actual notice of a particular problematic output being produced by its software. In such cases, companies can "implement a 'notice-and-blocking' system, loosely similar to 'notice-and-takedown' systems required under the [Digital Millennium Copyright Act] as to copyright and trademark infringements."[611]

Companies can engage in monitoring and oversight and, through extensive testing, mitigate the risks of autonomous AI algorithms.[612] Companies can fine-tune their algorithmic design models through

[605] *See* RONALD K.L. COLLINS & DAVID M. SKOVER, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE 27 (2018) (focusing on Siri's "limitations by design").

[606] *Id.*

[607] Burk, *supra* note 270, at 4 ("The OpenAI developers have put some safeguards on their device, so that it will decline to generate texts on certain controversial subjects (such as fascism or pedophilia), although some users have found ways to circumvent such safeguards."); *see id.* at 15 ("OpenAI has, for example, placed such "guardrails" on the ChatGPT system, so that it will usually decline to generate texts on certain topics, such as pedophilia or Nazism.").

[608] *See* Yan Luo, Xuezi Dan & Nicholas Shepard, *China Proposes Draft Measures to Regulate Generative AI*, GLOB. POL'Y WATCH (Apr. 17, 2023), https://www.globalpolicywatch.com/2023/04/china-proposes-draft-measures-to-regulate-generative-ai [https://perma.cc/9FFD-VVZA] (putting forth draft measures that would regulate generative AI in accordance with content moderation principles).

[609] *Id.*

[610] Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 28.

[611] *See* Volokh, *supra* note 563, at 493 (explaining that when AI companies have knowledge of a defamatory output of the system and avoid blocking it, the actual malice standard in defamation law may be satisfied).

[612] *See* Miriam Buiten, *Product Liability for Defective AI*, SSRN, July 19, 2023, at 1, 2, https://ssrn.com/abstract=4515202 [https://perma.cc/96K9-LP4J] (proposing a regime of product liability for algorithms).

design repairs and corrections.[613] Despite efforts to block recommendations, there can be errors, and it is up to regulators or courts to determine the scope of liability; they will have to outline the reasonable scope of programming, auditing, and monitoring sufficient for mitigating AI risks in terms of a cost-benefit analysis.[614] Given the human action behind the algorithms, there are strong justifications to hold companies accountable for damages caused by AI recommendation systems when companies fail to take reasonable steps to mitigate risks.[615]

2. *The problem of causal connection between targeting and harm*

The second potential difficulty of the proposed framework is the question of causal connection. First, how will courts know if a child's suicide or self-harm resulted from an algorithmic recommendation or was due to content the child sought out independently? A more important question is whether the child's suicide occurred because of the algorithmic targeting or because of the child's specific vulnerability. And what if the child was in a poor mental state and would have committed suicide anyway? One can assume that many children received algorithmic recommendations for the Netflix show *13 Reasons Why*,[616] but certainly not all of them committed suicide. Many children received a TikTok algorithmic recommendation for the "Blackout Challenge," but not all tried it.[617] As described earlier,[618] in the context of algorithmic recommendations that allegedly led to terror attacks, courts rejected lawsuits *inter alia* because of the question of causal connection.[619] In *Twitter, Inc. v. Taamneh*, the U.S. Supreme

[613] *See* Henderson et al., *supra* note 41, at 589 (suggesting that policymakers need to consider what technical design incentives they create).

[614] For expansion on cost-benefit analysis in the context of algorithmic recommendations, see Lavi, *Do Platforms Kill?*, *supra* note 44, at 525.

[615] Buiten, *supra* note 612, at 14–15 (explaining that courts might have to determine what failure algorithmic rate is reasonable).

[616] For more on this Netflix show, see *supra* note 2 and accompanying text.

[617] *See infra* note 646 and accompanying text (elaborating on the "Blackout Challenge" and its effects on children); *see also* Goldman, *Section 230*, *supra* note 373 (discussing the district court's disposition in *Anderson v. TikTok* and noting the lack of causal connection between the "Blackout Challenge" and the child's death).

[618] *See supra* Section II.B.1.d (overviewing court decisions addressing various platforms' liability for algorithms that incite terrorism).

[619] *See, e.g.*, Crosby v. Twitter, Inc., 921 F.3d 617, 624–26 (6th Cir. 2019) (rejecting the material support claim because of proximate cause requirements); Cain v. Twitter,

Court applied a proximate cause standard requiring a nexus between the defendant's assistance and the specific terror attack.[620] In other words, there must be a direct and specific connection between terrorism and the algorithmic recommendations.

How should the law interpret the causal connection requirement, and how can plaintiffs prove the presence of a causal connection between an algorithm recommendation and a child's self-harm? In the surveillance capitalism age, the bits of data transmitted through social media are countable,[621] and interactions and their frequency are measurable.[622] Social network analysis allows the measurement of transitions in social networks,[623] and transmitting a message is mappable.[624] Today, it is possible to determine whether a child was exposed to content via the algorithmic recommendation or if they reached it independently.

The question of causal connection can be complex. Yet, there are cases, such as *Maynard v. Snapchat*, where the court held that because the plaintiffs adequately alleged a causal connection between the harmful conduct and the technological design, on remand to the Court of Appeals, they were permitted to argue the trial court erred in dismissing their case for lack of proximate cause.[625] Moreover, in the surveillance capitalism age, establishing a causal connection is possible; Knight Professor of Constitutional Law and the First Amendment at *Yale Law School*, Jack M. Balkin, recently addressed the context of algorithmic amplification and explained that the causal connection between speech and harm seems increasingly measurable.[626] For example, regarding *13 Reasons Why*, experts warned Netflix that the show could promote teen suicide, *i.e.*, the company was put on notice, but Netflix ignored the warning, targeted teens in

Inc., No. 17-cv-02506-JD, 2018 WL 4657275, at *2 (N.D. Cal. Sept. 24, 2018) (dismissing the case due to lack of proximate cause)*;* Pennie v. Twitter, Inc., 281 F. Supp. 3d 874, 892 (N.D. Cal. Dec. 4, 2017) (finding no causal connection between Hamas and a Dallas shooter).

[620] Twitter, Inc. v. Taamneh, 598 U.S. 471, 505–06 (2023).

[621] Bambauer et al., *supra* note 95, at 519.

[622] *Id.*

[623] *Id.* at 533.

[624] Balkin, *Free Speech*, *supra* note 100, at 1250 (explaining that a central goal of the algorithmic society is the "collection and measurement of social phenomena" which "means converting human action, behavior, communication, and social life into measurable data, relationships, items, and commodities").

[625] Maynard v. Snapchat, 366 Ga. App. 507, 507–11 (2023).

[626] Balkin, *Free Speech*, *supra* note 100, at 1253.

particular, and only after robust criticism added a "viewer warning card" before the first episode.[627] Many children who committed suicide copied *13 Reasons Why* by leaving behind tapes with reasons for taking their lives in the same way as the main character of the show.[628]

Imposing liability for every exposure to harmful content via algorithmic recommendations would impose a heavy burden on social media, likely hindering its further development. Arguably, the causation elements could serve as the gatekeepers to litigation.[629] In some ways, the need to prove direct causation between algorithmic targeting and self-harm can protect social media companies as it will be hard for plaintiffs to prove,[630] leading courts to reject lawsuits regarding algorithmic targeting.[631] However, exempting intermediaries from liability is undesirable because it would lead to under-deterrence.[632] Thus, intermediaries would not have sufficient incentive to invest in efforts mitigating the risks of algorithmic design. We therefore need to re-conceptualize causation to strike a balance so that future development is not stifled *and* companies have sufficient incentive to invest in efforts that mitigate the risks of algorithmic design.[633] This cannot be achieved through an all-or-nothing approach to compensation but rather one based on probability.[634]

Intermediary liability for exposing children to harmful algorithmic targeting should be established when plaintiffs can prove a connection between the targeting and the harm.[635] If an intermediary breached its duty of care by indirectly targeting children with harmful recommendations, and children acted upon them, there should be a remedy, even if the causation was not direct. For example, if a child is

[627] *See supra* notes 2–10 and accompanying text.

[628] *E.g.*, Diaz, *supra* note 1.

[629] Cyphert & Martin, *supra* note 43, at 203 ("We believe that these causation elements—rather than Section 230—could serve as the gatekeeper to litigation.").

[630] *See id.* (discussing how courts should not throw out claims based on algorithmic amplification before the plaintiff reaches the discovery stage).

[631] *See* Henderson et al., *supra* note 41, at 634 (explaining that proving causation in an AI algorithmic society is difficult).

[632] *See* Stefan Heiss, *Towards Optimal Liability for Artificial Intelligence: Lessons from the European Union's Proposals of 2020*, 12 HASTINGS SCI. & TECH. L.J. 186, 205–06 (2021).

[633] *See id.* at 205 (explaining that "[n]ew emerging technologies can raise sophisticated causation issues").

[634] *Id.*

[635] *See* Lavi, *Do Platforms Kill?*, *supra* note 44, at 549–50 (proposing a defined legal duty of care regarding terrorists' unlawful content and that failure to meet this legal duty would strip companies of their immunity and allow for civil remedies).

targeted with pro-anorexia content and develops an eating disorder, the law should grant a remedy, even if the child could have been exposed to such content from other sources. However, in such cases, the compensation should be proportionate to the probability that the targeting caused the eating disorder.

Notably, even before the algorithmic society, many scholars argued that adherence to an all-or-nothing solution is inappropriate when there is a systematic infliction of harm and uncertainty in causation.[636] There are situations where the law should focus on the *increased risk*[637] or the *loss of chance* to survive.[638] Accordingly, courts should allow evidentiary lenience when the causation is uncertain because it is difficult for the plaintiff to prove that the defendant's wrongdoing caused the harm.[639] These doctrines, which provide a solution for the problem of uncertain causation, have been applied when there was information about other victims of the same tortfeasor who suffered similar harm.[640]

When many children who have suffered the same harm from targeting file a class action against a social media company, courts might be willing to impose liability under one of these doctrines,

[636] ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 125–29 (2001).

[637] *See* Howard Ross Feldman, Comment, *Chances as Protected Interests: Recovery for the Loss of a Chance and Increased Risk*, 17 BALT. L. REV. 139, 151 (1987) ("Increased risk situations arise where the negligent conduct of the defendant causes the plaintiff to incur an increased risk or susceptibility to a physical harm or disease."). Increased risk doctrine was adopted in relation to toxic torts, namely when victims were injured after exposure to dangerous substances caused harm. *See* Robert J. Rhee, *The Application of Finance Theory to Increased Risk Harms in Toxic Tort Litigation*, 23 VA. ENVTL. L.J. 111, 172–84 (2004). This doctrine was also applied in cases when conditions of employment "increased the risk to injury beyond that to which the general public was exposed." Matt Hlinak, *In Defense of the Increased-Risk Doctrine in Workers' Compensation*, 7 J. BUS. & ECON. RES. 57, 59 (2009). The loss of chances doctrine was mainly applied in cases of mass torts and medical malpractice that led to loss of chances to recover from an illness. *See* Benjamin Shmueli, *"I'm Not Half the Man I Used to Be": Exposure to Risk Without Bodily Harm in Anglo-American and Israeli Law*, 27 EMORY INT'L L. REV. 987, 998 (2013).

[638] Lavi, *Do Platforms Kill?*, *supra* note 44, at 560; Shmueli, *supra* note 637, at 998.

[639] *See* Benjamin Shmueli & Moshe Phux, *Small Data, Not (Only) Big Data: Personalized Law and Using Information from Previous Proceedings*, 35 OHIO ST. J. DISP. RESOL. 331, 363–64 (2020).

[640] *See id.* at 363 (arguing that courts can benefit from knowing about a defendant's repeat offenses in mass tortfeasor cases); Shmueli, *supra* note 637, at 999 (stating that the increased risk doctrine is often "presented in cases in which harm was eventually caused and the victim wishes to sue the party who increased the risk").

particularly the *loss of chances* doctrine.[641] The *increased risk* doctrine is less likely to apply because it refers to the increase in probability of future harm from targeting, which is overbroad and would likely result in flooding the courts with lawsuits.[642] Under this doctrine, every child who has been targeted with harmful content could file a lawsuit even if they never incurred actual harm.[643] In *Cohen v. Facebook, Inc.*, for example, the court dismissed a lawsuit for lack of standing because the plaintiff's case was based on increased risk for future harm, which was not "actual or imminent."[644] In contrast to the *increased risk* doctrine, the *loss of chances* doctrine focuses on harm that has already occurred and is better suited to cases of algorithmic targeting.[645]

For example, this doctrine should be applied to the "Blackout Challenge" case, which led to multiple deaths from strangulation among children who participated in the challenge.[646] It can be argued that the targeting exposed a susceptible population to dangerous content beyond what the general public is exposed to, thereby reducing the survival chances of the children who perished.

In such cases, courts would not grant damages in full, but rather, damages would be measured by multiplying the probability of occurrence. Compensation would be proportionate to the probability of *loss of chances*, even if the *loss of chances* falls below fifty percent.[647] Adopting a risk-based approach and rejecting an all-or-nothing approach to compensation overcomes the problem of uncertain causal connection and provides victims and their families with a remedy for the harm of targeting.[648]

[641] *See* Shmueli, *supra* note 637, at 999 (discussing the classic cases of loss of chance doctrine in which "both the damager and the extent of damage are known, but where it is not clear who among the group of victims suffered harm as a result of the actions of the damager and who was harmed due to some other factor").

[642] *See id.* (explaining the increased risk doctrine).

[643] *See id.*

[644] 252 F. Supp. 3d 140, 150 (E.D.N.Y. 2017); Clapper v. Amnesty Int'l, 568 U.S. 398, 401 (2013) (establishing that injury must be "certainly impending" to qualify for standing). Notably, after *Cohen v. Facebook*, courts continue to require concrete harm and not settle for the possibility for future harm. *See* TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2210 (2021).

[645] For a discussion of the differences between the two doctrines, see Lavi, *Do Platforms Kill?*, *supra* note 44, at 560–61.

[646] For an explanation of the "Blackout Challenge," see Anderson v. TikTok, Inc., 637 F. Supp. 3d 276, 278 (E.D. Pa. 2022).

[647] Lavi, *Do Platforms Kill?*, *supra* note 44, at 560 (arguing that compensation need not adhere to an all-or-nothing solution).

[648] *See id.* at 560–61 (discussing the potential benefits of adopting this approach).

3. Limitations on targeting: A First Amendment perspective

There is a distinction between freedom of expression and the First Amendment. The former is broader than the rights guaranteed by the latter, which "focuses on freedom of speech from government abridgment."[649] In the digital age, the First Amendment is less relevant to many aspects of online speech in light of the pluralist model shaping speech on online platforms and the global nature of online speech.[650] However, in the United States, companies use the First Amendment to avert regulation[651] and benefit from stronger freedom of speech protections than other democracies.[652] Unlike the EU, which balances free speech, privacy, and data protection, the United States limits speech only within narrowly defined categories without applying a proportionality test.[653] The next Sections will address First Amendment concerns regarding users' speech rights and, subsequently, corporate speech rights.

a. Users and the First Amendment

Arguably, regulation concerning limiting the targeting of speech and age verification obligations infringes users' free speech rights. The following Sections will explain why these arguments are without merit.

i. A user's First Amendment right to receive recommendations

One might contend that limiting algorithmic recommendations frustrates young users' right to access information and infringes their

[649] Balkin, *Free Speech*, *supra* note 100, at 1215, 1273 (explaining the pluralist model of speech regulation as one that has many different speech regulators that interpret the First Amendment differently).

[650] *Id.* at 1222 ("Because of the features of U.S. constitutional doctrine, when the free speech claims of end users and digital companies conflict, courts are likely to assign First Amendment rights to digital companies and not to end users.").

[651] *Id.* at 1210–11.

[652] Evelyn Douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 772 (2021); *e.g.*, Citizens United v. Fed. Election Comm'n, 558 U.S. 310, 342, 349, 353 (2010) (allowing corporations to make electioneering communications).

[653] Hadar Y. Jabotinsky & Michal Lavi, *The Eye in the Sky Delivers (and Influences) What You Buy*, 24 U. PA. J. CONST. L. 1329, 1394 (2022); *see* 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.").

First Amendment right to receive recommendations. There are many books and movies that include suicide, self-harm, or violence, and such content is protected by the First Amendment.[654] Importantly, the proposed framework does not challenge the publishing or exhibiting of such content; rather, it focuses on the algorithm's *targeting* of children. Considering the harm that could be caused by repeated recommendations targeted at children based on their previous activities and other knowledge companies have about users, such targeting breaches a duty of care.

Restricting the content targeted at children would not infringe on their right to receive information, as children would remain free to seek out this content using search engines. Indeed, the First Amendment protects the promotion of works to viewers who might "especially want to see them."[655] However, targeting children is not just promoting the works; instead, it is based on a trove of individualized data that can be used to manipulate and negatively influence children. Children may be bombarded with content they have not sought out, which can be very different from their initial choices of content.[656] These practices create a special context that warrants liability.

Modern free speech theory must take the potential harm of speech seriously, even beyond the context of influencing minors. Numerous laws already act to regulate speech in the workplace to address harassment and discrimination, balancing free expression and harm mitigation.[657] Similarly, governments seek to mitigate manipulative behaviors in commercial speech and prohibit "misleading and false" statements.[658]

In the case of targeting children, the "listener-centered" approach to regulation supports excluding children.[659] Targeting does not serve their free speech rights to information as listeners;[660] to the contrary,

[654] *See* Amici Curiae in Support of Netflix, *supra* note 99, at 4–10.

[655] *See id.* at 14.

[656] For example, if a child seeks content on dieting, the algorithm can bombard him with content on eating disorders and encourage anorexia or bulimia.

[657] *See* Cortez & Sage, *supra* note 266, at 760 (referring to Helen Norton, *Truth and Lies in the Workplace: Employer Speech and the First Amendment*, 101 MINN. L. REV. 31 (2016)).

[658] *Id.*; *see* Lavi, *Manipulating*, *supra* note 53, at 285.

[659] Norton, *Powerful Speakers*, *supra* note 274, at 454–55 (recognizing children are vulnerable listeners who should be protected from coercion).

[660] *Id.* at 467 (suggesting that the listener-centered approach balances the interests of the listeners and speaker but gives more weight to the listeners' interests).

corporate algorithmic targeting frustrates listeners' autonomy and self-governance. As explained, because of the special characteristics and vulnerabilities of children, personalized targeting can distort children's decision-making processes and inflict grave harm by using an opaque algorithm without transparency.[661] Moreover, because the algorithm repeatedly recommends content to children that they did not seek out wittingly, based on information collected on them, child users could feel so inundated by information fed to them that they no longer have the energy or desire to find content independently.[662] Hence, targeting could narrow the diversity of content to which children are exposed and curtail the marketplace of ideas.[663] Therefore, targeting children would not promote the First Amendment values of these susceptible listeners and could even frustrate these values. Thus, even assuming targeted content is protected speech, the interests of algorithmic "speakers"[664] and their susceptible listeners collide. Considering the special context of targeting children and the grave harm it can inflict, protecting young listeners should prevail.

ii. Age Verification and the First Amendment

Age verification is a barrier to users who want to visit a new website or application as it requires a short time delay before reaching the contents of the website or application.[665] This process is burdensome and might reduce users' willingness to consume or contribute content.[666] Further, users may hesitate to share personally identifying information to access content, particularly when the content is "sensitive or controversial."[667] This was the claim in the aforementioned case, *NetChoice, LLC v. Bonta*, which challenged the CAADCA.[668] However, unlike the CAADCA, which conditions access to certain websites, the age verification suggested in the proposed

[661] *See supra* Section I.D.

[662] *See supra* Section I.D.

[663] Lavi, *Targeting Exceptions*, *supra* note 77, at 133.

[664] For a normative analysis regarding algorithms free speech rights, see *supra* Section II.A.2.

[665] Goldman Amicus Brief, *supra* note 536, at 5.

[666] *Id.*

[667] *Id.* at 8.

[668] Assemb. B. 2273, 2022 Leg., Reg. Sess. (2024 Ca.); NetChoice, LLC v. Bonta, No. 22-CV-08861-BLF, 2023 WL 6135551, at *1 (N.D. Cal. Sept. 18, 2023).

framework would not limit access.[669] Age verification would only be required for users interested in algorithmic recommendations while others could freely use the website without verifying their age.[670]

Another possible argument is that requiring positive age verification infringes on the users' freedom of expression as it limits their anonymity and requires users to submit personal documentation; this, in turn, could chill their speech when they use the platform.[671] Free speech doctrines protect the speaker's right to conceal their identity, particularly where a speaker chooses anonymity in order to express unpopular or dissenting ideas.[672] Indeed, the right to anonymity is part of the freedom of speech; identification requirements tend to restrict the freedom to distribute information.[673] A series of cases has clarified that, in many circumstances, anonymity is a constitutional right.[674] Therefore, courts could strike down this regulation because of the right to communicate anonymously under U.S. law.[675]

However, the proposed framework would not obligate all users to verify their age, only users who want algorithmic recommendations. Users who want to avoid age verification could continue to use the platform without the recommendation system function. Moreover, the user's identity would not be exposed to all; it would only be revealed to the intermediary, which would use it only for age verification. The intermediary would be required to delete the identifying documents from the system upon verification or use safeguards, such as encryption or anonymization, to secure the information, and the data would not

[669] *See supra* notes 564–98 and accompanying text (explaining the proposed framework).

[670] *See supra* notes 564–98 and accompanying text (applying the framework to the respective circumstances).

[671] For a similar argument in a related context, see Jabotinsky & Lavi, *Speak Out*, *supra* note 589, at 577.

[672] JEFF KOSSEFF, THE UNITED STATES OF ANONYMOUS 50, 53–54 (2022); Lavi, *Manipulating*, *supra* note 53, at 335.

[673] KOSSEFF, *supra* note 672, at 42.

[674] *See* Jabotinsky & Lavi, *Speak Out*, *supra* note 589. *See generally*, Talley v. California, 362 U.S. 60 (1960) (voiding a Los Angeles City ordinance that forbade the distribution of any handbills if the handbills did not contain the name and address of the person by whom they were prepared); McIntyre v. Ohio Elections Comm'n, 514 U.S. 334 (1995) (voiding an Ohio statute prohibiting anonymous campaign literature and holding that such a law violates the First Amendment); Buckley v. Am. Const. L. Found., Inc., 525 U.S. 182, 198–200, 204 (1999); Watchtower Bible & Tract Soc'y of N.Y. v. Vill. of Stratton, 536 U.S. 127, 165–70 (2002).

[675] U.S. CONST. amend. I, cl. 2.

be used to unmask the user's identity. Thus, a substantial chill in users' speech is not expected.[676]

Another argument against the proposal could be that the very requirement for companies to verify the user's age limits the company's freedom to shape their system software; code is information, and information is a form of speech. Specific obligations to program a system to require *ex-ante* identity verification could be construed as infringing on the company's freedom of expression.[677] Thus, arguably, courts could strike down this obligation for violating the company's First Amendment rights. However, this type of technological code is not a core expressive interest warranting First Amendment protection.[678] The following Section will explain why arguments regarding corporate free speech[679] are also without merit.

b. A company's First Amendment right to engage in algorithmic targeting

Arguably, the recommendation system's targeting could be treated as constituting the speech of the company operating the platform.[680] Therefore, limitations on targeting children and other individuals who have not verified their age could infringe on the company's First Amendment rights because the proposed framework would include limitations on the targeting of both unprotected and protected speech. Some might contend that the *promotion* of speech is also protected alongside the *content* of the speech.[681]

Using this argument, companies could mobilize the First Amendment as a weapon for their own enrichment.[682] Scholars have warned that accepting such arguments could lead to a new Lochner Era, in which courts could invalidate the regulation of business based

[676] For a similar argument in the decentralized finance context, see Jabotinsky & Lavi, *Speak Out*, *supra* note 589, at 578–79.

[677] *Id.* at 579 (quoting Kyle Langvardt, *The Doctrinal Toll of "Information as Speech"*, 47 LOY. U. CHI. L.J. 761, 798–99 (2016)).

[678] *See* Jabotinsky & Lavi, *Speak Out*, *supra* note 589, at 580 (referring to 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, 502 (2020)).

[679] *See infra* notes 680–700 and accompanying text.

[680] *See* Section III.C.3.b (discussing how companies might argue that their recommendation system's targeting counts as speech).

[681] *See* Goldman Amicus Brief, *supra* note 536, at 14 (arguing that the First Amendment protects both the content and promotion of speech).

[682] Cheong, *supra* note 273, at 699.

on claims grounded in individual rights.[683] Such "modern corporate speech decisions rest on questionable theoretical grounds and make questionable assumptions"[684] and transfer the weight of First Amendment rights from political to commercial speech and from individuals to corporations.[685]

A counterargument is that First Amendment doctrine draws a line not only between protected and unprotected speech but also between speech and conduct.[686] As one scholar noted, "[I]f any online expression is considered speech, all online conduct, including ordinary commerce such as manufacturing, selling, buying, and contracting, would *absurdly* be subject to First Amendment protection."[687] Recommendation systems and targeting are business activities aiming to enhance the platform's profits and should not carry expressive values because algorithms do not "know" the recommended content or what the user data signifies. Accordingly, such AI outputs are non-speech in the first place and should not be constitutionally protected.[688] Targeting seeks to "maximize user engagement," which is industry code for including addictive features aimed at users, especially vulnerable populations such as children,[689] and such behavior, "[a]s any other commercial enterprise, [] can be regulated."[690]

[683] *Id.*; Cortez & Sage, *supra* note 266, at 714 ("Lochnerism may be particularly apt. The Lochner era refers to a forty-year period from 1897 to 1936 during which the Supreme Court struck down dozens of minimum-wage, labor, and other laws regulating business based on 'liberty of contract' and other un-enumerated economic rights.").

[684] Cortez & Sage, *supra* note 266, at 714.

[685] *See* MARY ANNE FRANKS, THE CULT OF THE CONSTITUTION 13 (2019) (discussing the shift in free speech priorities towards protecting commercial speech to the detriment of political speech).

[686] Citron & Franks, *supra* note 370, at 58.

[687] Cheong, *supra* note 273, at 700 (noting that "the internet is not 'a magical speech conversion machine,' so offline conduct not protected by the First Amendment should not be 'transformed into speech merely because it occurs online'").

[688] *See* MANHEIM & ATIK, *supra* note 270, at 4 ("All speech is communication but not all communication is speech. The First Amendment does not reach communication that is not speech.").

[689] *Id.* at 10–11.

[690] *Id.* at 11. Notably, this approach is reflected in a related context regarding the expressive value of content moderation. *See* NetChoice, LLC v. Paxton, 49 F.4th 439, 445 (5th Cir. 2022) (holding that platforms' decisions regarding content censorship are not protected under the First Amendment). The U.S. Supreme Court has taken up the case and will decide it in the near future. *See* Lawrence, *supra* note 86, at 30 (referring to *NetChoice, LLC v. Paxton*, 143 S. Ct. 744 (2023)).

Even if algorithmic targeting is beyond market behavior, arguably, it should be seen as machine speech, thereby garnering fewer rights than natural speech.[691] As explained, there is good reason to believe that algorithms should be attributed only secondary free speech rights, if any.[692] Therefore, protecting such speech is not at the core of the First Amendment because the potential chill focuses on algorithmic automated speech.[693]

Moreover, even if algorithmic targeting were to be treated as a type of protected speech, conceptualizing algorithm-generated content as speech should not automatically shield it against all regulation.[694] The protection accorded should not be core First Amendment protection; instead, due to the commercial nature and goals of targeting, regulation should subject it only to intermediate scrutiny standards[695] because there is a substantial governmental interest in protecting children from harmful targeting.[696] Targeting based on personalized information is not merely promoting content to the general public; rather, it is individualized and can be very harmful to children because of its selective repetition and exploitation of cognitive biases, which goes beyond content itself.[697] Therefore, prohibiting targeting of children is narrowly tailored considering its potential to cause tremendous harm.

The justification for extending First Amendment protection to commercial speech is based on the value of the information to its listeners and the marketplace of ideas.[698] However, as explained,[699] targeting children does not serve their free speech rights to

[691] *See, e.g.*, SIMON CHESTERMAN, WE, THE ROBOTS? 120 (2021); Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 46; Cortez & Sage, *supra* note 266, at 710.

[692] *See supra* note 277 and accompanying text.

[693] Jabotinsky & Lavi, *Can ChatGPT and the Like Be Your Co-Authors?*, *supra* note 79, at 46.

[694] *Id.*; Cheong, *supra* note 273, at 682–83.

[695] Lavi, *Manipulating*, *supra* note 53, at 331.

[696] *See supra* note 514 and accompanying text.

[697] Lavi, *Manipulating*, *supra* note 53, at 112.

[698] Norton, *Powerful Speakers*, *supra* note 274, at 443, 451; Norton, *Manipulation*, *supra* note 274, at 229; MANHEIM & ATIK, *supra* note 270, at 3 ("Algorithmic outputs in any particular instance may reflect the desires of the targeted 'listeners' more than the intention of any sender, autonomous or human.").

[699] Norton, *Powerful Speakers*, *supra* note 274, at 454–55 (recognizing children are vulnerable listeners who should be protected from coercion).

information as listeners; the argument of the value of speech to its listeners does not apply when targeting children.[700]

IV. TOWARDS A COMPREHENSIVE APPROACH TO ALGORITHMIC TARGETING

Broader solutions should be directed at the industry structure of surveillance capitalism and AI. Such regulation should focus on the design stage and platform operation instead of *ex-post* liability after the platform's operations have already caused harm.[701] As Professor Balkin explains, the origin of many of the digital public sphere's ills is the structure of the business model of surveillance capitalism, a characteristic of the algorithmic society.[702] Such models developed because companies have access to user data in the first place[703] and such access is without sufficient privacy protection, transparency, and oversight.[704] To mitigate the harm of targeting in general, Professor Balkin has asserted that there is a need to reform information capitalism and the algorithmic society.[705] This can be done first by focusing on privacy and data protection practices and second by focusing on the uses of AI.

As I have explained in a previous work,[706] one way to limit the harmful effects of personalized targeting is to focus on the upstream of the influencing stage by regulating the collection of information and data retention.[707] The EU General Data Protection Regulation[708] (GDPR) already regulates the collection and usage of data beyond the purpose of collection and retention of data;[709] it regulates the processing and retention of data in ways that limit the ability of data controllers to manipulate data subjects.[710] Such limitations could help

[700] *Id.* at 467.

[701] *See supra* note 563.

[702] Balkin, *Free Speech*, *supra* note 100, at 1243, 1269.

[703] *Id.* at 1243.

[704] *Id.* at 1269.

[705] *Id.* at 1270.

[706] Lavi, *Manipulating*, *supra* note 53, at 93.

[707] *Id.* at 313.

[708] General Data Protection Regulation, 2016 O.J. (L 119) 1.

[709] 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), art. 65, 2016 O.J. (L 119).

[710] General Data Protection Regulation, art. 17, 2016 O.J. (L 119) 1 (outlining limitations on the use of personal data).

reduce the influence of targeting and the extent to which individuals would be tied to their past digital activities.[711]

Even if the law does not take affirmative steps to enact specific legislation to regulate data collection and retention explicitly, the GDPR's influence already extends beyond the EU's borders as it applies to non-EU companies offering goods or services to EU consumers.[712] We may also be witnessing a broader "Brussels Effect" as data globalization has led U.S. legislators to enact data protection legislation in the United States.[713] The GDPR includes a criterion for assessing international transfers of personal data to countries beyond its borders and establishes a legal framework for preventing data exports to countries failing to meet this criterion.[714] The threshold for extraterritorial data transmission is the "adequacy" of data protection standards in the foreign jurisdiction.[715] Instead of relying on an adequacy determination, the EU and the United States have established the Privacy Shield, a voluntary compliance program within the private sector.[716] In order to align with the GDPR and facilitate data transfers between the EU and the United States, and given the absence of comprehensive data protection laws at the federal level in the United States, individual states are motivated to implement their own laws. For example, the California Consumer Privacy Act of 2018 (CCPA) applies to companies that do business in California.[717] It adopts some of the GDPR's key features, including a principle of data minimization, a right to access, and a right of consumers to know about data collection practices.[718] It also contains a right to erasure, ensuring that consumers have the right to request a business delete any personal

[711] *Id.* (requiring information controllers to comply with requests from data subjects to erase personal data).

[712] Lavi, *Manipulating*, *supra* note 53, at 317.

[713] *See* Anupam Chander, Margot E. Kaminski & William McGeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733, 1739 (2021) (referring to "Brussels Effect" as a race to the top in data protection standards).

[714] Paul M. Schwartz, *Global Data Privacy: The EU Way*, 94 N.Y.U. L. REV. 771, 783 (2019).

[715] *See* General Data Protection Regulation, 2016 O.J. (L 119) 1, 8.

[716] Lavi, *Manipulating*, *supra* note 53, at 317.

[717] CAL. CIV. CODE § 1798.140(d)(1); Lavi, *Manipulating*, *supra* note 53, at 318.

[718] CAL. CIVIL CODE §§ 1798.100(a)–(b), 1798.115, 1798.110, 1798.140(d), 1798.140 (v)(1)(L)(3)(y); Lavi, *Manipulating*, *supra* note 53, at 318–19.

information about them.[719] After the CCPA came into force, more states followed in its footsteps.[720]

Thus, even though the United States differs from the EU in its approach to free speech, some privacy protections can be, and already are, implemented in the United States and can make targeting less effective. As I have explained elsewhere, data retention makes targeting effective, as it shackles individuals to their past activities and decisions.[721] Therefore, limitations on data retention are a starting point for an overall reform of targeting.

U.S. policymakers are starting to realize the importance of imposing limitations on data retention. For example, the White House Office of Science and Technology Policy published a "Blueprint for an AI Bill of Rights"[722] that addresses, *inter alia*, the need for clear timelines for data retention.[723] Although this statement is not binding, it can impact and help mitigate the harm of targeting for everyone, not only for individuals who opt out of targeting altogether.[724]

A second broader solution that could mitigate the risk of targeting is regulation of AI risks. Because recommendation systems are based on AI algorithms, AI regulation can minimize harm.[725] However, the risks raised by AI are complex, and regulating processes can hinder technological innovation, especially if applied across sectors and before AI systems are brought to market.[726] Nevertheless, legislators aim to mitigate AI risks through regulation in both the EU and the United States.

[719] CAL. CIVIL CODE § 1798.105.

[720] *See, e.g.*, VA. CODE ANN. §§ 59.1-575–59.1-585 (West 2023); Colorado's Privacy Rights Act, COLO. REV. STAT § 6-1-1306 (2023). For further information, see Daniel J. Solove, *The Limitations of Privacy Rights*, 98 NOTRE DAME L. REV. 975, 983 (2023).

[721] Lavi, *Manipulating*, *supra* note 53, at 337.

[722] The White House, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf [https://perma.cc/UD8Y-9MUT].

[723] *See* Lavi, *Manipulating*, *supra* note 53, at 321 (referring to Keith E. Sonderling, Bradford J. Kelley & Lance Casimir, *The Promise and the Peril: Artificial Intelligence and Employment Discrimination*, 77 U. MIA. L. REV. 1, 41 (2022)).

[724] *Id.* at 322 (noting that the statement aims to mitigate manipulation by strengthening data privacy and safety of systems).

[725] *See* Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1369, 1372–73 (2023) (discussing risk regulation as a means of regulating AI systems).

[726] *See id.* at 1401 (explaining that AI systems are complex and complex systems are more likely to experience unpredictable and catastrophic risks).

In April 2021, the European Commission proposed the Artificial Intelligence Act ("AI Act").[727] This legislative proposal is set to become the world's first comprehensive legal framework for AI regulation, aiming to create a global standard for technology before implementation or expansion.[728] The AI Act is expected to come into effect two years after entry into force in 2026.[729] It focuses on two kinds of risks: "risks to health and safety and risks to fundamental rights."[730] The AI Act takes a risk management approach and divides AI system uses into four categories, subjecting each category of risk to different regulations[731]:

(1) *Unacceptable risks*: Prohibits certain applications, such as applications that comprise subliminal techniques[732] or systems that exploit vulnerabilities of a specific group of persons due to their age or physical or mental disability.[733]

[727] *Commission 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*]; *see* Kaminski, *supra* note 725, at 1374 ("The Draft EU AI Act aims to establish a comprehensive cross-sectoral European approach to governing AI systems. Once adopted and in effect, it will serve a harmonizing function, precluding EU Member States from enacting divergent laws—including more protective laws."); Solove, *Artificial Intelligence and Privacy*, *supra* note 121, at 21–23.

[728] Hanna Ziady, *Europe Is Leading the Race to Regulate AI. Here's What You Need to Know*, CNN BUS. (June 15, 2023, 11:19 AM), https://edition.cnn.com/2023/06/15/tech/ai-act-europe-key-takeaways/index.html [https://perma.cc/9PC3-25WT] ("Lawmakers have agreed a draft version of the Act, which will now be negotiated with the Council of the European Union and EU member states before becoming law."); Ana Hadnes Bruder, Oliver Yaros, Mark A. Prinsley, Rajesh De, Dominique Shelton Leipzig, Arsen Kourinian et al., *EU AI Act: European Parliament and Council Reach Agreement*, MAYER BROWN (Dec. 12, 2023), https://www.mayerbrown.com/en/perspectives-events/publications/2023/12/eu-ai-act-european-parliament-and-council-reach-agreement [https://perma.cc/VF4B-GQQN].

[729] Isabel Gottlieb, *EU AI Act's Passage Starts the Clock for US Companies to Comply*, BL (Mar. 13, 2024, 1:49 PM), https://news.bloomberglaw.com/artificial-intelligence/the-eu-parliament-just-voted-to-regulate-ai-what-happens-next [https://perma.cc/NTC3-APX3].

[730] Kaminski, *supra* note 725, at 1375.

[731] Claudio Calvino, *The Four Risks of the EU's Artificial Intelligence Act: Is Your Company Ready?*, FTI CONSULTING (July 25, 2023), https://www.fticonsulting.com/insights/fti-journal/four-risks-eus-artificial-intelligence-act [https://perma.cc/B8S6-743Z]. For further information on AI as a "risky complex system" reflected in the AI Act, see Kaminski & Jones, *supra* note 427, at 37.

[732] *Id.*

[733] *Artificial Intelligence Act*, *supra* note 727, at 43.

(2) *High risks*: Applications that pose a risk of harm to health and safety or an adverse impact on fundamental rights, such as biometric identification, are subject to detailed risk regulation.[734] These uses are also subjected to human oversight throughout their lifecycle,[735] a requirement for a human on the loop.[736] Additionally, high-risk AI systems should be "designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy, robustness, and cybersecurity, and perform consistently in those respects throughout their lifecycle."[737]

(3) *Limited risks*: AI systems that should be subject to transparency obligations.[738] For example, a duty to disclose that a person is engaging with a machine (*i.e.*, a bot) and thus allowing the person to make an educated decision whether to proceed with the conversation.[739]

(4) *Low or minimal risk*s[740]: Uses such as spam filters and inventory management systems already widely deployed[741] and encouraged by AI boards to follow voluntary self-regulation.[742]

Importantly, the AI Act is vague and does not allow flexibility in anticipating future risks and adapting to rapid technological advancements.[743] Despite its vagueness and shortcomings, the AI Act can prohibit or at least limit the targeting of children as it exploits their vulnerabilities and poses an unacceptable risk, or at least a high risk.

[734] Kaminski, *supra* note 725, at 1381 (noting that high-risk AI systems must comply with certain substantive and procedural requirements and undergo a "conformity assessment" before being placed on the European market).

[735] *Artificial Intelligence Act*, *supra* note 727, at 51.

[736] Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 504 (2023) ("The Act dictates that high-risk AI systems 'shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons.'").

[737] *Artificial Intelligence Act*, *supra* note 727, at 51–52.

[738] *Id.*

[739] *Id.* at 69; Calvino, *supra* note 731 ("For instance, an individual interacting with a chatbot must be informed that they are engaging with a machine so they can decide whether to proceed (or request to speak with a human instead).").

[740] *Artificial Intelligence Act*, *supra* note 727, at 12.

[741] Calvino, *supra* note 731.

[742] *Id.*

[743] *See* Michael von Lichtenstein, *Artificial Intelligence: Regulations vs. Innovation*, GIS (June 21, 2023), https://www.gisreportsonline.com/r/eu-ai [https://perma.cc/WRK8-W2B7]; *see also* Solove, *Artificial Intelligence and Privacy*, *supra* note 121, at 22 (explaining that it is unclear "who should" assess the harms and risks).

The U.S. approach to AI regulation, which also focuses on risk, has a lighter touch.[744] For example, the United States has the National Institute of Standards and Technology (NIST) and two proposed bills, the Algorithmic Accountability Act of 2022 and Washington State's SB 5116, which focus largely on impact assessment.[745]

The NIST's AI Risk Management Framework contains voluntary guidelines focusing on risk categorization, developed from the bottom up by multi-stakeholders.[746] The guidelines instruct organizations to map, measure, and manage risks throughout the process of coding and implementing AI,[747] ultimately leaving companies to decide both their risk tolerance and specific risk management practices.[748] A company's success in implementing such guidelines depends on its organizational culture for effective risk management and the implementation of accountability structures to challenge risky designs.[749] NIST reflects the values by design approach, which advocates for identifying and considering human needs and values in the design process.[750] The guidelines extend beyond *ex-ante* impact assessment, emphasizing risk management throughout the lifecycle of AI systems: the collection stage, training, and monitoring post-deployment.[751]

Alongside voluntary guidelines, some bills aim to regulate AI risks. The Algorithmic Accountability Act of 2022[752] mandates assessment and mitigation of AI risks[753] and requires the FTC to "create regulations and structures for companies to carry out assessments and provide transparency around the impact of automated decision-

[744] Kaminski, *supra* note 725, at 1379–80.

[745] *Id.* at 1351, 1374.

[746] *Id.* at 1374, 1377.

[747] *Id.* at 1383–84.

[748] *See id.* at 1384–85 (describing a company's duties under each step of the interactive process).

[749] *Id.* at 1385.

[750] Noëmi Manders-Huits & Jeroen van den Hoven, *The Need for a Value-Sensitive Design of Communication Infrastructures*, *in* EVALUATING NEW TECHNOLOGIES 51, 55 (Paul Sollie & Marcus Düwell eds., 1st ed. 2009); Deirdre K. Mulligan & Jenifer King, *Bridging the Gap Between Privacy and Design*, 14 U. PA. J. CONST. L. 989, 1019 (2012).

[751] Kaminski, *supra* note 730, at 1385.

[752] S. 3572, 117th Cong. (2022); H.R. 6580, 117th Cong. (2022).

[753] S. 3572, 117th Cong. (2022); H.R. 6580, 117th Cong. (2022).

making."[754] The Act aims to enable public accountability and use the impact assessment process as a policy-generating mechanism.[755]

A second prominent bill is the Washington State legislation SB 5116.[756] It focuses on *ex-ante* risks, outlining bans and standards regarding such risks.[757] The bill applies only to government actors and focuses on impact assessment;[758] however, it is more protective than other U.S. AI regulations as its risk assessment process resembles a licensing scheme.[759] Moreover, it has broad applicability, covering not just automated decisions with significant effects but any "procurement, development, and use of automated decision systems."[760] The bill includes regulatory tools beyond risk assessment, such as public registration of all automated decision systems used by public agencies and an obligation for an annual audit by the agency's Chief Information Officer.[761] Finally, the bill includes an individual right to be informed of the use of an automated system.[762]

There are alternative methods to reduce the risks associated with targeting besides civil liability in tort law. Regulation concerning the collection and retention of information could significantly reduce these risks.[763] Another option involves regulating AI; however, this approach might impede useful technology and innovation.[764] Further, soft risk regulation focusing on AI impact assessment tools might be insufficient to mitigate harm.[765] The law should therefore seek a sensitive balance between mitigation of harm and promotion of

[754] Section 5 of the Federal Trade Commission Act prohibits "unfair or deceptive acts or practices." *See, e.g.*, 15 U.S.C. § 45(a)(1); Ellen P. Goodman & Julia Trehu, *AI Audit Washing and Accountability*, SSRN, Sept. 22, 2022, at 16, https://ssrn.com/abstract=4227350 [https://perma.cc/NX3W-BABE] (draft); Kaminski, *supra* note 725, at 1381.

[755] The bill would be enforced by the FTC as part of its Section 5 authority against unfair practices. Kaminski, *supra* note 725, at 1380–81.

[756] *Id.* at 1374.

[757] *Id.* at 1381.

[758] *Id.* (highlighting that SB 5116 does not apply to private actors).

[759] *Id.*

[760] *See id.* at 1374, 1387 (noting that, for example, the GDPR "covers only solely automated systems that produce a significant effect on an individual").

[761] *Id.* at 1381.

[762] S.B. 5116, 67th Leg., Reg. Sess. § 4 (Wash. 2021).

[763] *See* Lavi, *Manipulating*, *supra* note 53, at 312.

[764] Michael von Lichtenstein, *supra* note 743.

[765] *See* Kaminski, *supra* note 725, at 1352, 1379–80 (referring to tools such as licensing, ongoing reporting, third-party audits, ongoing testing, and design mandates).

innovation by combining tools for *ex-ante* risk mitigation through platform design with process and *ex-post* liability.

CONCLUSION

Algorithms automatically target personalized recommendations without knowing the individuals they target because of the surveillance capitalism model and the AI's algorithmic capability to learn, predict, and perform tasks typically associated with human intelligence. Companies use these models to increase their traffic and enhance profits, and are indifferent to the consequences. Thus, as demonstrated,[766] they often target children with endless streams of harmful content that could result in detrimental behavior.

This Article addressed whether intermediaries operating social media platforms should bear liability for targeting susceptible children. It argues that the new ecosystem of surveillance capitalism and AI algorithmic recommendations have changed the environment in which information flows and transformed its context. With greater power must come greater responsibility. Laws must respond to these new models that allow companies to have a powerful influence on their targets, which is all the more true when the targets are children, as they are more impulsive and their decision-making abilities are underdeveloped.[767] Because of the new models and ecosystem and the unique susceptibility of children, a balance must be struck between normative considerations for liability; this new balance should consider that these models allow companies to wield powerful influence, especially over vulnerable populations.

In light of these new models, the normative rationales for free speech should be seen in a different light. This Article proposed acknowledging a duty of care regarding targeting children, which the law should treat as negligent design.[768] It proposed allowing children or their families to file civil lawsuits against companies that target harmful recommendations to children if the children who were bombarded with recommendations subsequently engaged in self-harm. I believe the proposed framework can mitigate the harm caused by targeting.

[766] *See supra* Introduction.

[767] *See supra* Section I.C.

[768] *See supra* Part III.

This Article concluded by addressing possible objections to the proposed framework,[769] such as the challenge of imposing liability on autonomous algorithms, the problem of causal connection and compensation, and First Amendment objections to liability. Finally, the Article explained that imposing *ex-post* liability on intermediaries for targeting children has limitations; therefore, a more comprehensive approach combining *ex-post* liability and *ex-ante* regulation of information and AI should be adopted.

[769] *See supra* Section III.C.