PLATFORM LIABILITY FOR PLATFORM MANIPULATION

*Sabriyya Pate**

Platform manipulation is a growing phenomenon affecting billions of internet users globally. Malicious actors leverage the functions and features of online platforms to deceive users, secure financial gain, inflict material harms, and erode the public's trust. Although social media companies benefit from a safe harbor for their content policies, no state or federal law clearly ascribes liability to platforms complicit in deception by their designs. Existing frameworks fail to accommodate for the unique role design choices play in enabling, amplifying, and monitoring platform manipulation. As a result, platform manipulation continues to grow with few meaningful legal avenues of recourse available to victims.

This Note introduces a paradigm of corporate liability for social media platforms that facilitate platform manipulation. It argues that courts must appreciate platform design as a dimension of corporate conduct by explicating the extension of common law tort liability to platform design. This Platform Design Negligence (PDN) paradigm crucially clarifies the bounds of accountability for the design choices of social media companies and is well-suited to respond to the law's systemic discounting of platform design. Existing legal frameworks fail to account for the unique and content-agnostic enmeshment between platforms and those who manipulate platforms to abuse users. PDN in turn offers a constitutive baseline for a society with less rampant technology-enabled deception.

* J.D. Candidate 2025, Columbia Law School. Many thanks to Professors David Pozen, Christopher Morten, and Raúl Carrillo for their invaluable guidance.

INTRODUCTION

Platform manipulation refers to the activity of malicious actors[1] who use social media platforms to deceive users.[2] It is implicated in a wide range of online activities—from online romance scams involving celebrity impersonators[3] to elder abuse whereby victims lose their life savings by "investing" with fraudsters.[4] Much to the chagrin of social media executives,[5] malicious actors identify and communicate with victims

1. For those who suspect that they are being targeted by a scam, know there are resources available for support. The AARP Fraud Watch Network Helpline is (877) 908-3360. A trained fraud specialist is available to provide free counseling and guidance between 8:00 AM and 8:00 PM ET, Monday through Friday.

2. "Manipulation" offers three meanings in the context of liability for social media companies. In this Note, "manipulation" in "platform manipulation" primarily refers to the practices of malicious actors, such as scammers, who exploit the design of platforms to achieve their desired outcomes. These manipulators largely seek to deceive platform users to secure financial gain. In this way, "platform manipulation" is a triple entendre; it refers to malicious actors' manipulation of the design of social media platforms, malicious actors' manipulation of social media users, and platforms' own manipulation of their users by way of their platform design.

3. See infra note 54.

4. See, e.g., Ann Pistone & Jason Knowles, Lombard Woman Loses Nearly $1 Million Life Savings in 'Pig Butchering' Scam, ABC7 Chi. (Sept. 4, 2024), https://abc7chicago.com/post/lombard-woman-loses-1-million-life-savings-pig-butchering-scam-forced-sell-home-belongings/15267382 [https://perma.cc/5LZP-XCTQ].

5. When pressed on the widespread romance scams on his platform, the then-Match Group Chief Executive Officer replied, "[T]hings happen in life." Jim Axelrod, Sheena

through reputable social media platforms like Facebook, Instagram, and Match.com, as well as non-social media platforms like Amazon and Cash App.[6] In doing so, these actors exploit the functions and features that make online platforms attractive digital spaces to begin with.

Platform manipulation creates irreparable harm to individuals from all walks of life. For starters, it creates tremendous financial harm. Platform manipulation is part of a booming multibillion-dollar industry in the United States.[7] In 2022, fraudsters stole over $137 billion from Americans,[8] and those over age sixty lose approximately $28.3 billion from scams each year.[9] Successful scams that involve "deepfakes," such as artificial intelligence (AI)-generated nude images of minors, can also create long-lasting reputational and psychological harm to victims.[10] In some

---

Samu, Andy Bast & Matthew Mosk, As Romance Scammers Turn Dating Apps Into "Hunting Grounds," Critics Look to Match Group to Do More, CBS News (Apr. 24, 2024), https://www.cbsnews.com/news/romance-scams-dating-apps-investigators-match-group [https://perma.cc/DJ2U-FC63] (internal quotation marks omitted) (quoting Bernard Kim) (describing the death of Laura Kowal after she matched with a scammer on Match.com).

6. See Edward C. Baig, 8 Warning Flags to Help You Find Fraudulent Apps, AARP (Sept. 10, 2021), https://www.aarp.org/home-family/personal-technology/info-2021/warning-signs-of-fraudulent-apps.html [https://perma.cc/75C4-9BT4] (last updated Feb. 13, 2024) ("Nearly 2 percent of the 1,000 highest-grossing apps on the App Store are scams . . . .").

7. See Emma Fletcher, Social Media: A Golden Goose for Scammers, FTC (Oct. 6, 2023), https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2023/10/social-media-golden-goose-scammers [https://perma.cc/NL5T-SFW7] [hereinafter Fletcher, Golden Goose] ("Scammers are hiding in plain sight on social media platforms and reports to the FTC's Consumer Sentinel Network point to huge profits."). Today, more than half of Americans have a friend or family member who has been scammed, and Americans receive approximately thirty-three million robocalls each day. Alana Semuels, The Government Finally Did Something About Robocalls, TIME Mag. (Dec. 15, 2023), https://time.com/6513036/robocalls-government-action/ [https://perma.cc/2H58-YR4X]; Survey: Most Americans Know Someone Targeted by Scam, ABA Banking J. (Nov. 15, 2024), https://bankingjournal.aba.com/2024/11/survey-most-americans-know-someone-targeted-by-scam/ [https://perma.cc/G5UU-8STL].

8. FTC, Protecting Older Consumers 2022–2023, at 40 (2023), https://www.ftc.gov/system/files/ftc_gov/pdf/p144400olderadultsreportoct2023.pdf [https://perma.cc/8EC3-4AFW].

9. Michael Rubinkam, Scammers Are Swiping Billions From Americans Every Year. Worse, Most Crooks Are Getting Away With It, AP News (July 7, 2024), https://apnews.com/article/scammers-billions-elder-fraud-aarp-ai-f9530303e10b998720414e88430bcf6b (on file with the *Columbia Law Review*) (citing Jilenne Gunther, The Scope of Elder Financial Exploitation: What It Costs Victims, AARP BankSafe Initiative 1 (2023), https://www.aarp.org/content/dam/aarp/money/scams-and-fraud/2023/true-cost-elder-financial-exploitation.doi.10.26419-2Fppi.00194.001.pdf [https://perma.cc/U93E-UJLP]).

10. See Dana Nickel, AI Is Shockingly Good at Making Fake Nudes—And Causing Havoc in Schools, Politico (May 29, 2024), https://www.politico.com/news/2024/05/28/ai-deepfake-nudes-schools-states-00160183 (on file with the *Columbia Law Review*) ("Students in New Jersey, Florida, California and Washington state have reported embarrassing deepfake experiences that can result in arrests or nothing at all, a gap in laws that can leave victims feeling unprotected.").

instances, victims have attempted to rob banks for their scammers.[11] One man in Ohio killed an Uber driver who he wrongfully suspected of involvement with a scam.[12] At a meta level, platform manipulation poses many implications for a global society: Democratic discourse necessitates the kind of trust that online scammers extract from public spheres.[13]

Platform manipulators rely on the core fabric of social media platforms—their user interfaces (UI) and user experiences (UX)—to operationalize and scale their exploitation.[14] These actors use platforms to identify and initiate communication with their targets.[15] They also leverage platforms to expand their operations, test new tactics, and hone their craft, often flying under the radar of platforms' content detection systems.[16]

Platform designs take many forms and can serve discrete goals. For example, platforms make design choices on how to display features; hiding the "reply all" feature can reduce accidental mass replies, while hiding the number of digits in passcodes can provide additional security. Though some social media companies have adopted platform designs that mitigate harms like cyberbullying and misinformation,[17] broadly, social media

---

11. See 74-Year-Old Ohio Woman Charged in Armed Robbery of Credit Union Was Scam Victim, Family Says, AP News (Apr. 24, 2024), https://apnews.com/article/ohio-credit-union-robbery-scam-arrest-23fe2c0a7f839d23c8796f04313ca522 (on file with the *Columbia Law Review*) (stating that relatives of a seventy-four-year-old woman claimed she was an online scam victim who was driven to commit armed robbery in order to "solve her financial problems").

12. See Ben Finley, What We Know About the Shooting of an Uber Driver in Ohio and the Scam Surrounding It, AP News (Apr. 19, 2024), https://apnews.com/article/uber-driver-killed-scam-4998a42b2e59aed3dda95f983b2f9b52 (on file with the *Columbia Law Review*) (describing how an Ohio man "fatally shot an Uber driver" because he mistakenly believed she was part of a scheme to extort $12,000 dollars, though she was also a scam victim sent by scammers to the shooter's house to pick up a supposed package).

13. See Evelyn Douek, Content Moderation as Systems Thinking, 136 Harv. L. Rev. 526, 540 (2022) (describing the impact of trust and transparency on social media consumers).

14. What Is the Difference Between UI and UX?, Figma, https://www.figma.com/resource-library/difference-between-ui-and-ux/ [https://perma.cc/9E22-33FW] (last visited Jan. 24, 2025) (describing user interface as the "interactivity, look, and feel of a product . . . while user experience (UX) covers a user's overall experience with the product or website").

15. See, e.g., Cordelia Lynch, SCAM: Inside Asia's Criminal Network, Sky News (Oct. 18, 2024), https://news.sky.com/story/they-fall-in-love-with-me-inside-the-fraud-factories-driving-the-online-scam-boom-13234505 [https://perma.cc/2HLV-X5W4] ("Based in highly secretive, heavily guarded compounds, fraud factories—similar to the ones Poom-Jai worked in—are spread across South East Asia, where the online scam industry has exploded.").

16. See, e.g., Isabelle Qian, 7 Months Inside an Online Scam Labor Camp, N.Y. Times (Dec. 17, 2023), https://www.nytimes.com/interactive/2023/12/17/world/asia/myanmar-cyber-scam.htm (on file with the *Columbia Law Review*) ("The workers spent their days using the WeChat accounts, swiping over social media feeds on each device to mimic normal use and get past the app's fraud detection system.").

17. See Amer Owaida, Instagram Rolls Out New Features to Help Prevent Cyberbullying, We Live Sec. (Apr. 23, 2021), https://www.welivesecurity.com/2021/04/23/

companies offer limited features to address scams and other kinds of platform-based deception.[18]

Meanwhile, it is exceedingly difficult for scam victims to get in touch with customer service personnel who could be positioned to assist them.[19] Payment provider platforms used by malicious actors to receive money from victims have been woefully unable to curb this problem, which often originates on social media platforms.[20] In recognition of the complexities of platform manipulation, some companies have begun to initiate voluntary commitments to "shar[e] insights and knowledge about the lifecycle of scams" with the goal of educating users on what to look out for.[21] While these efforts are positive developments, they at best indicate a growing recognition that social media companies lack direction when looking to design their platforms in ways that limit harm caused by the

---

instagram-new-features-curb-cyberbullying/ [https://perma.cc/T8ZF-XRLB] (explaining Instagram's new "abusive Direct Messages" filter and "a tool to stop someone a user has blocked from contacting them from another account," both designed to combat cyberbullying and abusive behavior on the platform).

18. See Kristina Radivojevic, Christopher McAleer, Catrell Conley, Cormac Kennedy & Paul Brenner, Social Media Bot Policies: Evaluating Passive and Active Enforcement 5 (Sept. 27, 2024) (unpublished manuscript), https://arxiv.org/pdf/2409.18931 [https://perma.cc/SS43-R7L9] (testing the social media platforms Facebook, Instagram, LinkedIn, Mastodon, Reddit, Threads, TikTok, and X and finding that all fail to sufficiently identify and respond to platform manipulation).

19. See, e.g., Steven John & Alexander Johnson, How to Contact Facebook Support and Get Help for Issues With Your Account, Bus. Insider (Sept. 19, 2023), https://www.businessinsider.com/guides/tech/how-to-contact-facebook-problems-with-account-other-issues [https://perma.cc/L6R7-BEBL] ("Don't bother trying to call Facebook.").

20. Social media companies are thus the "first responders" for many scams and fraudulent activities. Federal agencies have already sued major banking platforms, such as Zelle, for "for failing to protect consumers from widespread fraud." Laurel Wamsley, In a Lawsuit, CFPB Says 3 Top U.S. Banks Failed to Protect Consumers From Zelle Fraud, Or. Pub. Broad. (Dec. 24, 2024), https://www.opb.org/article/2024/12/24/cfpb-alleges-3-banks-failed-to-protect-consumers-from-zelle-fraud/ [https://perma.cc/T7J7-UWAU] (internal quotation marks omitted) (quoting Press Release, CFPB, CFPB Sues JPMorgan Chase, Bank of America, and Wells Fargo for Allowing Fraud to Fester on Zelle (Dec. 20, 2024), https://www.consumerfinance.gov/about-us/newsroom/cfpb-sues-jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle [https://perma.cc/9698-XPSB]).

21. See, e.g., Announcing the Tech Against Scams Coalition, Coinbase (May 21, 2024), https://www.coinbase.com/blog/announcing-the-tech-against-scams-coalition (on file with the *Columbia Law Review*) ("This partnership aims to protect and educate users, emphasizing that scams are a tech-wide issue, not limited to social media, crypto, or finance." (emphasis omitted)); Press Release, Aspen Inst., Aspen Institute Financial Security Program Launches National Task Force for Fraud & Scam Prevention (July 18, 2024), https://www.aspeninstitute.org/news/task-force-on-fraud-and-scams [https://perma.cc/UQC2-RSRX] ("The task force formalizes a network of stakeholders who have a vested interest in making sure that consumers are protected and can restore trust in our financial system.").

ballooning scam economy.[22] At worst, social media companies' short-term profit incentives directly converge with those of the malicious actors on their platforms.[23] It is also worth noting that social media users are better able to participate in the economy and generate advertising revenue when their funds are not siphoned into scammers' accounts.

As major platforms cobble together written policies to address platform manipulation,[24] companies face few legal restrictions on the

---

22. See Heather Kelly, The Nonstop Scam Economy Is Costing Us More Than Just Money, Wash. Post (July 13, 2022), https://www.washingtonpost.com/technology/2022/07/13/scam-fraud-fatigue/ (on file with the *Columbia Law Review*) ("Constant scam attempts can increase stress levels and strain relationships. Their negative impact on mental health is even worse when the scammers target people based on perceived weaknesses, like advanced age, loneliness or[,] . . . an ongoing illness.").

23. Social media companies profit off users' engagement on their platforms, including engagement with scammers. This engagement is packaged and sold to data brokers and advertisers. See Kalev Leetaru, What Does It Mean for Social Media Platforms to "Sell" Our Data?, Forbes (Dec. 15, 2018), https://www.forbes.com/sites/kalevleetaru/2018/12/15/what-does-it-mean-for-social-media-platforms-to-sell-our-data/ (on file with the *Columbia Law Review*) (describing how social media companies profit by selling user data to data brokers, developers, and advertisers).

24. See, e.g., Andrew Hutchinson, Meta Highlights Key Platform Manipulation Trends in Latest 'Adversarial Threat Report', Soc. Media Today (Nov. 30, 2023), https://www.socialmediatoday.com/news/meta-platform-manipulation-trends-adversarial-threat/701230/ [https://perma.cc/QGL7-VYUE] [hereinafter Hutchinson, Meta Highlights Key Platform Manipulation Trends] (discussing Meta's Q3 2023 "Adversarial Threat Report"); Community Standards, Meta, https://transparency.meta.com/policies/community-standards/ (on file with the *Columbia Law Review*) (last visited Jan. 24, 2025) ("Meta recognizes how important it is for Facebook, Instagram, Messenger and Threads to be places where people feel empowered to communicate, and we take our role seriously in keeping abuse off the service. That's why we developed standards for what is and isn't allowed on these services."); Countering Influence Operations, TikTok, https://www.tiktok.com/transparency/en-us/countering-influence-operations/ (on file with the *Columbia Law Review*) (last visited Jan. 24, 2025) ("This post explains how we continuously work to detect and disrupt covert influence operations that try to undermine the integrity of our platform, so that millions can continue to enjoy a safe, creative, and trusted TikTok experience."); Fake Engagement Policy, Google, https://support.google.com/youtube/answer/3399767?hl=en [https://perma.cc/RX74-NVYV] (last visited Jan. 24, 2025) ("YouTube doesn't allow anything that artificially increases the number of views, likes, comments, or other metrics either by using automatic systems or serving up videos to unsuspecting viewers. Also, content that solely exists to incentivize viewers for engagement (views, likes, comments, etc[.]) is prohibited."); How Does YouTube Address Misinformation, YouTube, https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/ [https://perma.cc/YE84-RXHL] (last visited Jan. 24, 2025) ("YouTube does not allow misleading or deceptive content that poses a serious risk of egregious harm."); How We Prevent the Spread of False Information on Snapchat, Snap (Sept. 8, 2022), https://values.snap.com/news/how-we-prevent-the-spread-of-false-information-on-snapchat [https://perma.cc/3XUU-F9PF] ("[Snapchat's] policies have long prohibited the spread of false information."); Misinformation, Meta, https://transparency.fb.com/policies/community-standards/misinformation (on file with the *Columbia Law Review*) (last visited Jan. 24, 2025) (explaining Meta's policies against misinformation); Platform Manipulation and Spam Policy, X (Mar. 2023), https://web.archive.org/web/20231216113944/https://help.twitter.com/en/rules-and-

design choices that render their platforms attractive breeding grounds for scammers.[25] In the absence of binding legal obligations on social media companies, malicious actors are free to play platforms like instruments of manipulation.

Existing legal frameworks constitute a patchwork of schemes that provide state and federal enforcers and citizens few chances to have their injuries heard, let alone to vindicate their rights and pursue remedies.[26] Innovative litigation strategies, such as the application of false advertising claims by private plaintiffs and the Federal Trade Commission (FTC), are stopgap solutions that have not steadied the problem.[27] The cornerstone of social media law, Section 230 of the Communications Decency Act of 1996, as well as First Amendment law and consumer law frameworks, all either fail to provide recourse to social media scam victims or fail to explain legislative inaction in the face of the causal relationship between platforms' design choices and the scams that transpire on those very same platforms.[28] Furthermore, maladaptation of § 230's immunity for platforms has created an inaccurate presumption of immunity for all choices, including design choices, made by social media companies.[29]

This Note is the first to argue for a social media liability paradigm that centers platform design choices: a Platform Design Negligence (PDN) paradigm that establishes the circumstances for a clear assumption of liability in this digital environment. It offers a roadmap for an evolution in law and society towards coherent parlance on the impacts of twenty-first century platform technologies. Social media companies should face liability when their design choices contribute to the deception of their users. When companies are aware of these deception risks and fail to take reasonable precautions, they cease to function as reasonable platforms and should become liable for injuries that follow. Through a full-throated

---

policies/platform-manipulation (on file with the *Columbia Law Review*) ("We want X to be a place where people can make human connections, find reliable information, and express themselves freely and safely. To make that possible, we do not allow spam or other types of platform manipulation.").

25. See Caleb N. Griffin, Systematically Important Platforms, 107 Cornell L. Rev. 445, 514 (2022) ("[C]ompanies that utilize manipulative technologies have no clear corporate law duties to rein in their behavior and protect their users from exploitation and other harms.").

26. See id. at 489–99 (discussing various state and federal regulatory efforts and noting that "few proposed regulations have successfully been made into law, and those few that are operative apply only in narrow contexts").

27. See, e.g., Forrest v. Meta Platforms, Inc., 737 F. Supp. 3d 808, 820–21 (N.D. Cal. 2024) (involving a man whose Facebook profile was used by scammers to create fake profiles); Press Release, L.A. Cnty. Dist. Att'y's Off., NGL Labs Charged in Consumer Protection Lawsuit (July 9, 2024), https://da.lacounty.gov/media/news/ngl-labs-charged-consumer-protection-lawsuit [https://perma.cc/KT5H-LHQW] (involving a social messaging app that deceptively marketed its platform to users); see also infra section II.B (describing U.S. consumer law's systemic discounting of social media platform users' rights).

28. See infra section II.B.

29. See infra section II.B; infra notes 236–239.

adoption of this paradigm, victims and law enforcers could hold social media companies accountable for harms caused by manipulation conducted on, by, and through their platforms. Both federal and state courts, without the mandate of a statute, can actualize this paradigm by applying and building upon existing common law tort doctrine.[30]

In Part I, this Note surveys the landscape of platform manipulation, discussing the harms caused by platform-based deception as well as the design choices that enable platform manipulation in practice. It also explores how social media companies profit from the scam economy. Part II turns to the absence of legal frameworks that apply to social media companies' design choices in the context of platform manipulation. It underscores the relationship between platform design and platform manipulation. It also delineates the pitfalls of the prevailing voluntary self-governance paradigm for platform manipulation. Finally, Part III introduces the PDN paradigm that can serve social media companies, lawmakers, and victims as they pursue legal remedies and design interventions that curb the growing challenge of platform manipulation.

I. PLATFORM MANIPULATION AND EXISTING FRAMEWORKS

Platform manipulation is a type of activity on social media[31] whereby malicious actors use platforms to manipulate users.[32] Platform manipulators are inherently rulebreakers: bad faith actors logged onto social media to purposefully manipulate social media users. Platform manipulation, as all forms of manipulation, is difficult to police due to the complexity of the dignity and autonomy rights at issue.[33] Yet many if not all social media companies are attuned to platform manipulation. For example, X defines platform manipulation as interactions with the social media platform that are done to "mislead others and/or disrupt their experience by engaging in bulk, aggressive, or deceptive activity."[34]

---

30. See infra Part III.

31. See Michael S. Rosenwald, Before Twitter and Facebook, There Was Morse Code: Remembering Social Media's True Inventor, Wash. Post (May 24, 2017), https://www.washingtonpost.com/news/retropolis/wp/2017/05/24/before-there-was-twitter-there-was-morse-code-remembering-social-medias-true-inventor/ (on file with the *Columbia Law Review*) (describing the genesis of contemporary platform-based social media).

32. For one example of a discussion of platform manipulation within platform governance legal scholarship, see generally Daphne Keller, Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard, 1 J. Free Speech L. 227 (2021) (explaining the difficulties in regulating platform manipulation).

33. Id. at 265; see also Cass R. Sunstein, Fifty Shades of Manipulation, 1 J. Mktg. Behav. 213, 219 (2015) (addressing why manipulation is rarely addressed both legally and politically).

34. Platform Manipulation, X (July 28, 2022), https://transparency.x.com/en/reports/platform-manipulation#2021-jul-dec [https://perma.cc/7W55-3WF5] (defining X's platform manipulation policy and highlighting a 2% global increase in "global anti-spam challenges" and a 6% increase in "global spam reports" since its last reporting period).

Platform manipulators use an array of tactics and maintain several objectives.[35] Those tactics include "social media bots," or coordinated fake accounts that aim to influence opinions.[36] Bots can pose as "real" users from one country and prolifically post propaganda praising or defending the actions of a different country, with the objective of portraying global support for a particular political posture.[37] Platform manipulators may also use social media to "giv[e] a false impression that there is genuine grass-roots support or opposition for a particular group or policy."[38] This is often referred to as misinformation or disinformation, depending on its intent.[39] Such platform manipulation schemes have contributed to real-world violence,[40] led ordinary people to attend and participate in manufactured in-person protests,[41] and more.[42] Most commonly, however, platform

35. See, e.g., Tim Wu, Is the First Amendment Obsolete?, 117 Mich. L. Rev. 547, 565–68 (2018) (describing disinformation campaigns as a form of "reverse" censorship that drowns out the truth or accurate depictions).

36. See Hutchinson, Meta Highlights Key Platform Manipulation Trends, supra note 24 (describing Meta's efforts to take down accounts that "aimed to sway discussion around both U.S. and China policy by both sharing news stories, and engaging with posts related to specific issues").

37. Id.

38. Franziska Keller, David Schoch, Sebastian Stier & JungHwan Yang, It's Not Easy to Spot Disinformation on Twitter. Here's What We Learned From 8 Political 'Astroturfing' Campaigns., Wash. Post (Oct. 28, 2019), https://www.washingtonpost.com/politics/2019/10/28/its-not-easy-spot-disinformation-twitter-heres-what-we-learned-political-astroturfing-campaigns/ (on file with the *Columbia Law Review*) (explaining the operations of social media disinformation campaigns).

39. Misinformation campaigns involve the dissemination of false information, regardless of intention to deceive, whereas disinformation campaigns involve the dissemination of misleading or biased information with the intent to manipulate. See Dean Jackson, How Disinformation Impacts Politics and Publics, Nat'l Endowment for Democracy, https://www.ned.org/wp-content/uploads/2018/06/How-Disinformation-Impacts-Politics-and-Publics.pdf [https://perma.cc/85K2-B5LH] (last visited Jan. 24, 2025) ("In the long-term, disinformation can be part of a strategy to shape the information environment in which individuals, governments, and other actors form beliefs and make decisions.").

40. See id. (discussing the communal violence sparked by the spread of false claims in India).

41. See id. (discussing manufactured protests in Germany).

42. One notable example of platform manipulation was the case of a Russia-linked company that posted content—posing as American users—aimed at driving wedges within the ideological spectrum in advance of the 2016 and 2020 U.S. presidential elections. See Young Mie Kim, New Evidence Shows How Russia's Election Interference Has Gotten More Brazen, Brennan Ctr. for Just. (Mar. 5, 2020), https://www.brennancenter.org/our-work/analysis-opinion/new-evidence-shows-how-russias-election-interference-has-gotten-more [https://perma.cc/ZB89-GZME] (finding "that social media accounts linked to the Internet Research Agency (IRA), the Kremlin-linked company behind an influence campaign that targeted the 2016 elections, have indeed already begun their digital campaign to interfere in the 2020 presidential election").

manipulation takes place in monotonous direct and group messaging features—hidden from public view.[43]

Platform manipulation is rooted in a centuries-old practice: "[C]ommon and 'enduring psychological [consumer] vulnerabilities' and 'cognitive and emotional susceptibilities' have forced 'industrialized and industrializing societies on every continent . . . [to] confront[] . . . commercial misrepresentation.'"[44] As human behaviors and cognition have changed in relation to social media,[45] the sophistication of consumer scams has similarly evolved. The FTC has reported on new and sophisticated dark patterns designed to deceive consumers.[46] Industry experts have identified troubling trends in the scam industry,[47] and a survey of fraud and risk professionals found widespread concern over the applications of AI to create even more complex scams.[48]

Platform manipulation is a difficult problem to define in the legal liability context because of the challenges with discerning the actors, intentions, and potential chilling effects of enforcement.[49] These

---

43. Accurate reporting on the scale of social media scams is difficult to ascertain given the conflict of interest. See supra note 23. Publicly available reports indicate that the cost of these scams is in the billions. See Fletcher, Golden Goose, supra note 7 ("Reported losses to scams on social media [between 2021 and October 2023] hit a staggering $2.7 billion, far higher than any other method of contact.").

44. David Adam Friedman, Imposter Scams, 54 U. Mich. J.L. Reform 611, 616 (2021) [hereinafter Friedman, Imposter Scams] (second, third, fourth, fifth, and sixth alterations in original) (quoting Edward Balleisen, Fraud: An American History From Barnum to Madoff 5 (2017)).

45. See Chantal Line Carpentier, UN Economist Network, New Economics for Sustainable Development: Attention Economy 1 (2025), https://www.un.org/sites/un2.un.org/files/attention_economy_feb.pdf [https://perma.cc/R66T-MPH2] ("To address the scarcity of people's attention, these technologies have been increasingly aimed at strategic capture of private attention aided by systematic collection and analysis of personal data, which has become a very profitable business model.").

46. See Press Release, FTC, FTC Report Shows Rise in Sophisticated Dark Patterns Designed to Trick and Trap Consumers (Sept. 15, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/09/ftc-report-shows-rise-sophisticated-dark-patterns-designed-trick-trap-consumers [https://perma.cc/LVM4-UTJ6] ("As more commerce has moved online, dark patterns have grown in scale and sophistication, allowing companies to develop complex analytical techniques, collect more personal data, and experiment with dark patterns to exploit the most effective ones.").

47. See Quinn Owen, How AI Can Fuel Financial Scams Online, According to Industry Experts, ABC News (Oct. 11, 2023), https://abcnews.go.com/Technology/ai-fuel-financial-scams-online-industry-experts/story?id=103732051 [https://perma.cc/MWB7-5MMJ] (discussing how "[g]enerative AI tools can make scams faster and more sophisticated").

48. Id. ("There is growing fraud online in which scammers manufacture other identities to dupe financial institutions or their customers out of money—and the crimes are only expected to grow more frequent with the increasing prevalence of artificial intelligence, experts say.").

49. See Jason Pielemeier, Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?, 2020 Utah L. Rev. 917, 923–26 (describing difficulties in

conceptual challenges carry over to platform operations, as platforms must first define platform manipulation in order to act upon it. In individual instances of platform manipulation, it is hard for social media companies "to objectively establish and measure harm."[50] Additionally, because "individuals or entities . . . targeted for enforcement . . . will often be able to justifiably complain about selective enforcement," platforms are incentivized to avoid taking adverse actions against their users, including platform manipulators.[51]

A. *Platform Manipulation Harms*

Platform manipulation consistently creates financial, reputational, psychological, and other harms for victims and their communities. Similar to victims harmed by poorly designed car safety systems or exercise equipment, those affected by platform manipulation carry a burden into their lives for extended periods.[52]

1. *Financial Effects.* — Platform manipulation is predicated on a requisite degree of human manipulation, and malicious actors frequently manipulate unsuspecting consumers for financial gain. Success is contingent on a scammer's ability to understand the "target's" personality, affectation, motivations, and desires.[53] There is a wide range of scam types, including phishing scams, romance scams,[54] impersonation scams, and

---

ascertaining "blame" in large-scale disinformation efforts due to inauthentic dissemination and organic amplification, as well as the potential for chilling economic activity).

50. Id. at 923–24.

51. Id. at 924; see also Mike Isaac & Theodore Schleifer, Meta Says It Will End Its Fact-Checking Program on Social Media Posts, N.Y. Times (Jan. 7, 2025), https://www.nytimes.com/live/2025/01/07/business/meta-fact-checking (on file with the *Columbia Law Review*) (last updated Feb. 3, 2025) ("Social media companies are increasingly relying on fact-checks written by their users, allowing companies to step back from politically loaded decisions about what content to take down.").

52. See Martina Barash, Toyota Settles Hybrid Owners' Individual Brake-Defect Claims, Bloomberg L. (Mar. 29, 2022), https://news.bloomberglaw.com/product-liability-and-toxics-law/toyota-settles-hybrid-owners-individual-brake-defect-claims (on file with the *Columbia Law Review*) (describing the resolution of a case in which car owners experienced several car crashes as a result of a brake defect); Sacramento Kings Reach Settlement With Sporting Goods Companies in Francisco Garcia Case, Sports Litig. Alert (Nov. 16, 2012), https://sportslitigationalert.com/sacramento-kings-reach-settlement-with-sporting-goods-companies-in-francisco-garcia-case/ [https://perma.cc/L6SQ-LLLG] (describing the resolution of a case in which an athlete suffered "significant injuries" due to use of gym equipment sold without a warning describing its risks).

53. See Kristy Holtfreter, Michael D. Reisig & Travis C. Pratt, Low Self-Control, Routine Activities, and Fraud Victimization, 46 Criminology 189, 209 (2008) (finding that self-control and remote purchasing play a role in fraud victimization).

54. See Jeannine Mancini, A Woman Loses $50,000 Thinking Elon Musk Was Telling Her 'I Love You' and Wanted to Make Her Rich—But It Was an Elaborate Deepfake Scam, Yahoo Fin. (Apr. 29, 2024), https://finance.yahoo.com/news/woman-loses-50-000-thinking-155924192.html (on file with the *Columbia Law Review*) (describing how a South Korean woman was scammed into sending $50,000 dollars to an Elon Musk impersonator through a romance scam that originated on Instagram).

even foreclosure relief scams.[55] For example, some scams have targeted student loan borrowers on social media; the scammers offer fake debt relief payment programs and pocket the entire amounts intended as student loan payments.[56] Scammers operating on platforms such as Indeed, ZipRecruiter, and Facebook can garner trust and exploit consumers through convoluted schemes that ask for money in return for hypothetical jobs.[57]

Americans lose billions of dollars from scams that are facilitated on social media.[58] Often times platform manipulators engage in a practice known as "pig butchering," in which users are "fatten[ed]"—or coerced into making greater contributions—over time before the ultimate "slaughter" leaves the victim penniless.[59] Such operations targeting social media users are global and complex.[60] In Myanmar, a single criminal network used "an army of modern-day slaves" to scam social media

---

55. What Are Some Common Types of Scams?, CFPB, https://www.consumerfinance.gov/ask-cfpb/what-are-some-common-types-of-scams-en-2092/ [https://perma.cc/427C-CZKF] (last visited Jan. 24, 2025) (defining blackmail scams, charity scams, debt collection scams, foreclosure relief scams, grandparent scams, impostor scams, lottery or prize scams, money mule scams, and sale-of-nonexistent-goods scams); Scam Glossary, FCC, https://www.fcc.gov/scam-glossary [https://perma.cc/9Z34-6EDE] (last visited Jan. 24, 2024) (providing a comprehensive glossary of scams with links to resources for all types of scams). For an example of a recent scam, see Meghan Bragg, A New Scam Is Making the Rounds on Facebook. How to Spot It: VERIFY, WCNC Charlotte (Aug. 15, 2023), https://www.wcnc.com/article/news/verify/verify-facebook-scam-warning-red-flags-to-avoid-becoming-victim/275-0cff86b5-a453-45a9-8cd2-02308bd51074 [https://perma.cc/R7X5-2PSB] (describing account verification scams on Facebook).

56. See Annie Nova, Student Loan Borrowers Should Be Aware of Debt Relief Scams, CNBC (Nov. 29, 2023), https://www.cnbc.com/2023/11/29/student-loan-borrowers-should-be-aware-of-debt-relief-scams.html [https://perma.cc/383C-Z9BW] ("Some scammers may falsely claim to be affiliated with the Department of Education or your servicer. Borrowers should be extra careful that they're actually speaking to someone at their servicer and might want to ask to call their lender back directly if they're having doubts."); Warning: Student Debt Relief Scam Circulating on Social Media, Charter Coll., https://chartercollege.edu/news-hub/warning-student-debt-relief-scam-circulating-social-media/ [https://perma.cc/8EDE-ZUGF] (last visited Jan. 24, 2025) (describing a student debt relief scam).

57. See What Are Some Common Types of Scams, supra note 55 ("Money mules may be recruited through online job or social media posts that promise easy money for little effort.")

58. See Fletcher, Golden Goose, supra note 7 (noting reports of nearly $3 billion in social media scam losses reported to the FTC in a two-year period).

59. See Cezary Podkul, What's a Pig Butchering Scam? Here's How to Avoid Falling Victim to One., ProPublica (Sept. 19, 2022), https://www.propublica.org/article/whats-a-pig-butchering-scam-heres-how-to-avoid-falling-victim-to-one [https://perma.cc/HMX6-XHH5].

60. See, e.g., Qian, supra note 16 ("Increasingly, people from India, the Philippines and more than a dozen other countries have also been trafficked [due to pig butchering] to work for scam gangs, prompting Interpol to declare the trend a global security threat.").

consumers out of $1 billion from their life savings.[61] Scammers do not tend to discriminate when choosing their targets.[62] Men and women, young people and elderly people, citizens and immigrants, among many others, are all targeted.[63]

2. *Reputational Effects.* — Platform manipulation, particularly disinformation and misinformation, creates reputational harm to victims, from the most vulnerable children to the highest-profile politicians, journalists, and celebrities. Teenagers and convicted predators alike have used AI-deepfake technology to manufacture nude images of individuals, including children.[64] Deepfakes can even convince people that their political leaders are dead.[65] Once these manipulated images are

61. See Teele Rebane, Ivan Watson, Tom Booth, Carlotta Dotto, Marco Chacon & Mark Oliver, Billion-Dollar Scam, CNN (Dec. 27, 2023), https://edition.cnn.com/interactive/2023/12/asia/chinese-scam-operations-american-victims-intl-hnk-dst/ (on file with the *Columbia Law Review*) (highlighting one scam operation "assembled by what the UN has called one of the largest human trafficking events in Asia in recent history").

62. A look at platform-enabled consumer scams like these serves the function of "looking to the bottom," in which victims are most disconnected from the social media companies and lawmakers in positions to address the problem. See Mari J. Matsuda, Looking to the Bottom: Critical Legal Studies and Reparations, 22 Harv. C.R.-C.L. L. Rev. 323, 324–25 (1987) (explaining the need to adopt the perspective of the least advantaged).

63. See, e.g., Juan Manuel Pedroza, Anne Schaufele, Viviana Jimenez, Melissa Garcia Carrillo & Dennise Onchi-Molin, Insurgent Citizenship: How Consumer Complaints on Immigration Scams Inform Justice and Prevention Efforts, 37 Geo. Immigr. L.J. 369, 372 (2023) (describing the range of scams targeting noncitizens in the U.S. and the obstacles faced by noncitizen victims of immigration scams); Anthony Hill, In-Depth: Top Scams That Are Targeted Against the Black Community; How to Avoid Falling Victim, ABC Action News (Aug. 12, 2021), https://www.abcactionnews.com/news/in-depth/in-depth-top-scams-that-are-targeted-against-the-black-community-how-to-avoid-falling-victim [https://perma.cc/CP2T-7ZN8] (stating that government imposter scams are more common within the Black community); Tom Huddleston Jr., Americans Are Being Scammed Out of Billions on Social Media—Look for These 7 Red Flags, CNBC (Oct. 12, 2023), https://www.cnbc.com/2023/10/12/americans-lose-billions-to-social-media-scams-red-flags-to-spot.html [https://perma.cc/MGL9-TV7V] (last updated Nov. 14, 2023) (describing how "[y]ounger [social media] users are especially at risk" for scams because they may be "overly trusting of the technology they're using" (internal quotation marks omitted) (quoting David McClellan, CEO, Soc. Catfish)); Matthew Rodriguez, Fake ICE Agent Indicted for Offering Green Cards to Undocumented Immigrants, CBS News (May 25, 2023), https://www.cbsnews.com/losangeles/news/fake-ice-agent-indicted-for-offering-green-cards-to-undocumented-immigrants/ [https://perma.cc/VJ5Z-KA85] (describing an ICE agent impersonator who charged up to $20,000 for immigration services).

64. See Lexi Lonas Cochran, From Deepfake Nudes to Incriminating Audio, School Bullying is Going AI, The Hill (June 6, 2024), https://thehill.com/homenews/education/4703396-deepfake-nudes-school-bullying-ai-cyberbullying/mlite/ (on file with the *Columbia Law Review*) (describing how teenagers have weaponized deepfakes against their classmates).

65. See Ali Swenson & Christine Fernando, As Social Media Guardrails Fade and AI Deepfakes Go Mainstream, Experts Warn of Impact on Elections, PBS News (Dec. 27, 2023), https://www.pbs.org/newshour/politics/as-social-media-guardrails-fade-and-ai-deepfakes-go-mainstream-experts-warn-of-impact-on-elections (on file with the *Columbia Law Review*)

introduced onto the internet, it becomes impossible to easily delete content that may have been downloaded or shared across platforms. This content can also be forwarded to traffickers, pedophiles, and others who abuse individuals offline.[66]

Platform manipulation also creates reputational harm in the traditional sense—victims who are deceived through social media are highly unlikely to report platform-enabled consumer scams due to embarrassment and shame.[67] As almost 40% of Americans do not understand that gullibility is not the cause of victimization, perceived reputational harms are amplified by general lack of information on the form and function of these scams.[68]

3. *Psychological Effects.* — When individuals are deceived through social media, there is also a mental and emotional component to the harm. In the most tragic cases, scam victims lose their lives. Ryan Last, a high-achieving high school student, succumbed to a "sextortion" scam in which a romance scammer solicited an explicit image of Last.[69] The scammers repeatedly asked Last for more money and added more pressure.[70] Last later died by suicide, leaving a note that detailed the embarrassment he felt for himself and his family.[71] Psychological consequences of scams include fear, shame, difficulty forming trusting relationships, difficulty engaging in digital interactions altogether, depression, anxiety, post-traumatic stress disorder, and other behavioral changes.[72] Platform

(explaining how deepfakes showing a president being rushed to a hospital could "spread without labels and fool people days before an election").

66. See Charles Toutant, An AI Took Her Clothes Off. Now a New Lawsuit Will Test Rules for Deepfake Porn, N.J. L.J. (Feb. 5, 2024), https://www.law.com/njlawjournal/2024/02/05/an-ai-took-her-clothes-off-now-a-new-lawsuit-will-test-rules-for-deepfake-porn/ [https://perma.cc/AH49-R4D3] (describing how photos from an Instagram page can be downloaded and manipulated into a doctored nude image).

67. Christina Ianzito, Many Victims Struggle With Mental Health in Scams' Aftermath, AARP (Dec. 15, 2022), https://www.aarp.org/money/scams-fraud/mental-health-impact/ [https://perma.cc/Q3V4-LL7Z] (explaining the negative mental health consequences faced by scam victims).

68. See Press Release, AARP, AARP Report: Americans Agree That Fraud is at a Crisis Level (May 17, 2023), https://press.aarp.org/2023-5-17-AARP-Report-Americans-Agree-Fraud-is-at-Crisis-Level [https://perma.cc/VX5Y-9CHS] ("Fraud is a severely under-reported crime, even as nearly nine in 10 adults feel people should report incidents. Nearly 40% of Americans still don't understand that victims do not lose money to scams because they are gullible. Victimization from a scam can happen to anyone.").

69. Josh Campbell & Jason Kravarik, Teen Boy's Death Hours After Scam Is Part of Troubling Increase in 'Sextortion' Cases, FBI Says, ABC 7 Chi. (May 21, 2022), https://abc7chicago.com/ryan-last-death-san-jose-ca-sextortion-scam/11877764/ [https://perma.cc/RZ98-P4KZ].

70. Id.

71. See id. ("'He really, truly thought in that time that there wasn't a way to get by if those pictures were actually posted online,' [Ryan's mother] Pauline said. 'His note showed he was absolutely terrified. No child should have to be that scared.'").

72. The Psychological Impact of Being Scammed: Safeguarding and Healing in the Digital Age, Sec. Everywhere (Dec. 28, 2023), https://www.security-everywhere.com/the-

manipulation is also a vector for race- and gender-based discrimination and harassment because actors are able to exploit platform designs to propagate harmful ideologies and target users based on their identities.[73]

International criminal networks rely on platform manipulation to commit direct physical violence as well. These criminal networks have been known to post fake jobs to recruit individuals to show up at distant locations; once they arrive, the scammers force the now-human trafficking victims to work at scam centers where they must pay off their "debt" through cybercrime.[74] In this way, platform manipulation schemes can psychologically damage both the victims and the perpetrators of online scams.

B. *Platform Design in Practice*

Platform design refers to the choices made to create the visual experience of interacting with platforms. This is often referred to as UI or UX design.[75] Platforms functionally facilitate introductions between scammers and their targets, and they recommend scammer content to consumers.[76] Platforms also play an important role in monitoring the prevalence of these scams, including by choosing how to design and implement "reporting flows" for such activity on their platforms.[77]

---

psychological-impact-of-being-scammed-safeguarding-and-healing-in-the-digital-age/ [https://perma.cc/D64B-NCJJ].

73. See, e.g., Spencer Overton & Catherine Powell, The Implications of Section 230 for Black Communities, 66 Wm. & Mary L. Rev. 107, 127–41 (2024) (describing how platforms facilitate anti-Black harassment and intimidation, "create online havens for white supremacists," enable advertisers to promote discriminatory services, and spread election misinformation that targets Black voters).

74. Juliana Kim, Online Scamming Industry Includes More Human Trafficking Victims, Interpol Says, NPR (Dec. 10, 2023), https://www.npr.org/2023/12/10/1218401565/online-scamming-human-trafficking-interpol [https://perma.cc/F27Q-ZLUA].

75. See Hany Farid & Brandie M. Nonnecke, The Case for Regulating Platform Design, Wired (Mar. 13, 2023), https://www.wired.com/story/make-platforms-safer-regulate-design-section-230-gonzalez-google/ [https://perma.cc/JNJ5-H2JP] ("Holding platforms accountable for negligent design choices that encourage and monetize the creation and proliferation of harmful content is the key to addressing many of the dangers that persist online.").

76. See Rohit Chopra & Samuel A.A. Levine, The Case for Resurrecting the FTC Act's Penalty Offense Authority, 170 U. Pa. L. Rev. 71, 117–18 (2021) ("[P]latforms earn almost all of their revenue by building detailed dossiers on users that can then be deployed to target advertising to individual consumers. . . . 'Targeted' or 'behavioral' advertising raises a host of consumer protection and competition concerns, including privacy, discrimination, fraud, and unfair competition." (footnotes omitted)).

77. See Twitter's New Reporting Process Centers on a Human-First Design, X (Dec. 7, 2021), https://blog.twitter.com/common-thread/en/topics/stories/2021/twitters-new-reporting-process-centers-on-a-human-first-design (on file with the *Columbia Law Review*).

Social media companies admit to struggling to design platforms in ways that dampen pervasive platform manipulation.[78] In turn, design choices about the interfaces that direct individuals to separate websites or downloads can play a major role in enabling social media scams.[79] Moreover, social media companies design their platforms to retain users.[80] They complicate reporting so that scam victims are not able to seek help from the platforms.[81] They fail to deploy labels and alerts in ways that could nudge victims and hinder scammers.[82] And information about these harmful platform designs is often buried in Terms of Service (ToS) agreements that are systematically unfair, imbalanced, and coercive.[83]

While some platforms have deployed "pre-bunking" measures,[84] major social media companies have not created dedicated scam prevention teams that rival their anti–political misinformation teams for platform manipulation more broadly.[85] As such, scam victims may receive limited assistance when engaging in drawn-out conversations with scammers that the platforms are privy to.[86]

Social media companies similarly fail to design UIs that provide embedded and aptly timed information on their policies. For example, while securities enforcement laws govern the practices of financial advisors on social media and fraudulent financial services are "prohibited" by platforms themselves, platforms are still hotbeds for investment-related scams, and the law is evolving to neglect the role of platform design in securities fraudsters' schemes to defraud.[87] The Financial Industry

---

78. See Coinbase, supra note 21 (describing scams as "a pervasive issue across the entire tech landscape" and a "challenge" that "requires a collective effort").

79. Such "drive-by downloads" account for 48% of cyberattacks on platforms. Michael McGuire, Social Media Platforms and the Cybercrime Economy 2 (2019), https://itcafe.hu/dl/cnt/2019-02/151108/bromium.pdf [https://perma.cc/8QJG-2GFP].

80. See infra section I.B.1.

81. See infra section I.B.2.

82. See infra section I.B.4.

83. See Michael L. Rustad & Thomas H. Koenig, Wolves of the World Wide Web: Reforming Social Networks' Contracting Practices, 49 Wake Forest L. Rev. 1431, 1436 (2014) (asserting that ToS are "systematically unfair and imbalanced" and proposing reforms "to expand the readability and standardiz[ation]" of disclosures).

84. See infra section I.B.5.

85. See Shannon Bond, False Information Is Everywhere. 'Pre-Bunking' Tries to Head It off Early, NPR (Oct. 28, 2022), https://www.npr.org/2022/10/28/1132021770/false-information-is-everywhere-pre-bunking-tries-to-head-it-off-early (on file with the *Columbia Law Review*) (describing efforts by Google and Twitter to test "pre-bunking," a strategy that "show[s] people the tactics and tropes of misleading information before they encounter it in the wild—so they're better equipped to recognize and resist it").

86. Cf. Lizzie O'Leary, Meta's Laid-Back Approach to User Hacking, Slate (Jan. 29, 2023), https://slate.com/technology/2023/01/instagram-facebook-meta-hacking-customer-support.html [https://perma.cc/98H7-XWT2] (describing difficulties with getting in touch with Meta customer support when user accounts are hacked).

87. See FINRA Staff, Investor Alert: Social Media "Investment Group" Imposter Scams on the Rise, Yahoo Fin. (Jan. 17, 2024), https://finance.yahoo.com/news/investor-

Regulatory Authority (FINRA) has even sent out an investor alert about actors posing as "registered investment advisors" that claim to be brokers and steal billions from consumers.[88] Social media companies nonetheless fail both to enforce their policies and to display the relevant terms anywhere near the areas where these scams are promoted.[89] By designing their UX to obfuscate the ToS,[90] social media platforms can readily gain ill-informed user consent, while pervasive mandatory arbitration clauses within agreements further preclude user action in response to deceptive design practices.[91]

Moreover, in the United States, social media companies are not required to use meaningful age-verification procedures, let alone profile-verification procedures.[92] As a result, malicious actors can create a universe of fake friends or followers that can create a strong impression that a fake account is real and allow scammers to scale their operations.[93] For

---

alert-social-media-investment-100000532.html (on file with the *Columbia Law Review*); Prohibited Financial Products and Services, Meta, https://transparency.meta.com/policies/ad-standards/deceptive-content/prohibited-financial-products-and-services (on file with the *Columbia Law Review*) (last visited Jan. 24, 2025) ("Advertisers can't run ads for financial products and services that are frequently associated with misleading or deceptive promotional practices.").

88. FINRA Staff, supra note 87.

89. See id. ("FINRA has seen a recent significant spike in investor complaints resulting from recommendations made by fraudulent 'investment groups' promoted through social media channels."); see also Przemysław Pałka, Terms of Service of Social Media Platforms, *in* Research Handbook on Social Media and the Law (Thaddeus Hoffmeister & Marilyn Bromberg eds., forthcoming 2025) (manuscript at 20) ("Put simply: it is in the platforms' direct interest to 'addict' people to their services. Further, it is in their interest for the law not to notice or regulate the potential externalities of such an addiction." (footnote omitted)).

90. See Johnathan Yerby & Ian Vaughn, Deliberately Confusing Language in Terms of Service and Privacy Policy Agreements, 23 Issues Info. Sys. 138, 146 (2022) (describing how social media platforms confuse or hide policies and controls from users).

91. See Kavya Jha & Ananya Singh, The Use of Arbitration Clauses by Social Media Websites: A Critique, 23 Pepp. Disp. Resol. L.J. 303, 306–09 (2023) (explaining how 40% of notable social media platforms have mandatory arbitration clauses); Caroline Marshall & Sarah Reynolds, Schillings, With 'Legal But Harmful' Gone, Will Terms of Service Protect Social Media Users?, Lexology (Feb. 23, 2023), https://www.lexology.com/library/detail.aspx?g=856c1650-3680-4ac7-87d1-d0f7b9cb47ac [https://perma.cc/WUW3-7YB3] (describing the lack of transparency around ToS and challenges with ToS being written vaguely); Cadie Thompson, What You Really Sign Up for When You Use Social Media, CNBC (May 20, 2015), https://www.cnbc.com/2015/05/20/what-you-really-sign-up-for-when-you-use-social-media.html [https://perma.cc/U7TX-QUVU] (last updated May 27, 2015) ("Social media giants not only have a license to use content that you post, but they are also constantly collecting data on you that you may not realize you are sharing.").

92. See Andrew Chung & John Kruzel, US Supreme Court Grapples With Texas Online Porn Age-Verification Law, Reuters (Jan. 15, 2025), https://www.reuters.com/legal/texas-online-porn-age-verification-law-goes-us-supreme-court-2025-01-15/ (on file with the *Columbia Law Review*) (describing the forthcoming Supreme Court case in which the Roberts Court is expected to rule on, among other items, whether online age verification "stifles the free speech rights of adults").

93. See infra section I.B.7.

example, in July 2024, Meta removed over 63,000 accounts on its platform that were operating sextortion scams; one network of 20 criminals was operating 2,500 fake accounts.[94] Celebrity imposter scams—in which scammers create accounts that purport to be well-known figures, develop relationships over social media, and then use those relationships to extort money—similarly rely on social media platforms permitting duplicate fake accounts that share the same names, photos, and other details.[95] One study of platform manipulation tactics has found that scammers commonly share accounts that are used to communicate with victims.[96] Scammers can also evade scam-detection mechanisms by "utilizing visually similar symbols to obfuscate their text, abusing account names, and splitting text into multiple comments posted by multiple accounts."[97]

Some platform design choices that bear less heavily but still significantly on platform manipulation include the "infinite scroll," the decision to allow consumers to view metrics (i.e., "likes" and "retweets") directly on posts, the decision to make all new accounts public by default, and the decision to impose word or character limits.[98] Aware of platform manipulation at its present scope, social media companies have various tools at their disposal when designing platforms in ways that are more (or less) conducive to deceptive conduct. Not all these platform design choices are presently permissible under the prevailing platform liability paradigm.[99] These capacities are inherent to the genesis of platform-based

---

94. Olivia Carville, Meta Removes 63,000 Accounts Linked to Sextortion Scammers, Bloomberg (July 24, 2024), https://www.bloomberg.com/news/articles/2024-07-24/meta-removes-63-000-accounts-linked-to-sextortion-scammers (on file with the *Columbia Law Review*).

95. See, e.g., 'National Geographic' Photographer Paul Nicklen Warns About Social Media Impostors, Part 2, AARP (Dec. 15, 2023), https://www.aarp.org/podcasts/the-perfect-scam/info-2023/paul-nicklen-part-2.html [https://perma.cc/UA8N-XYSP] ("Paul Nicklen is a world-famous wildlife photographer with a massive Instagram following . . . . He faces a near constant stream of impostors and he just can't seem to get social media companies interested in fixing the problem.").

96. See Xigao Li, Amir Rahmati & Nick Nikiforakis, Like, Comment, Get Scammed: Characterizing Comment Scams on Media Platforms 12 (2024) (unpublished manuscript), https://www.ndss-symposium.org/wp-content/uploads/2024-60-paper.pdf [https://perma.cc/V3KL-GHYH] (describing indicators that scammers share account details, that "multiple scammers [exist] inside a single campaign," and that those scammers exchange information about targets' identities).

97. Id. at 1.

98. See, e.g., Dayna Tortorici, Infinite Scroll: Life Under Instagram, The Guardian (Jan. 31, 2020), https://www.theguardian.com/technology/2020/jan/31/infinite-scroll-life-under-instagram [https://perma.cc/Z87Y-PLEE] (offering one account of the impacts of the infinite scroll design feature).

99. Often, these design elements are subject to "A/B testing" to "track the effect of design changes" and ultimately increase user "time on the platform." Maya Konstantino, Note, The Tort of Moving Fast and Breaking Things: A/B Testing's Crucial Role in Social Media Litigation, 99 N.Y.U. L. Rev. Online Features 178, 189–90, 202 (2024), https://nyulawreview.org/wp-content/uploads/2024/08/99-NYU-LRev-Online-178-1.pdf

digital technologies, and companies can leverage them to satisfy their burden of responsibility to users.

1. *Retention Features.* — Features that incentivize social media users, both scammers and victims, to continue to engage in dangerous activities on platforms are one potential avenue for ascribing liability for platform manipulation. "Retention features" include the infinite scroll, reward systems for repeat or sustained use of platforms, and other features that make it easier to conduct exchanges of money, images, or content.[100] The Ninth Circuit has recognized liability for a retention feature when a Snapchat filter allegedly encouraged dangerous driving.[101] Such features involve choices that only the platform and the individual user are privy to.[102] Consequently, platforms could deploy different retention features for different demographics, all the way down to the individual user basis. For example, platform designs that abandon the infinite scroll feature could limit the unique toll the infinite scroll takes on individuals who are predisposed to fraud online: those with poor mental health or memory.[103] Through retention design, platforms make active choices to retain users, including those who violate their policies and manipulate others on the platform."

2. *"Flows."* — Social media companies design their platforms in ways that affect usability and accessibility, and thus platform manipulation, through the number and sequencing of steps required in order for a user to effect a change to their UX. For example, a "reporting flow" refers to the steps required for a user to report an account for suspected deceptive activity: By designing more streamlined ways to submit and visualize reports on the user end, social media companies can create intuitive

---

[https://perma.cc/Y99B-398P] (discussing applications of a product liability framework to social media platforms).

100. See id. at 214 ("[TikTok] capitalize[s] on reward-based learning, infinite scroll, videos that consume the entire screen, and algorithmic manipulation, among other factors.").

101. See Lemmon v. Snap, Inc., 995 F.3d 1085, 1092 (9th Cir. 2021) (describing how "the duty that Snap allegedly violated 'springs from' its distinct capacity as a product designer" (quoting Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1107 (9th Cir. 2009))).

102. See id.; Kathleen Walch, How Generative AI Is Driving Hyperpersonalization, Forbes (July 15, 2024), https://www.forbes.com/sites/kathleenwalch/2024/07/15/how-generative-ai-is-driving-hyperpersonalization/ [https://perma.cc/N8YJ-HYD6] ("The idea of these uniquely personalized experiences is transforming how businesses interact with their customers and how people are living their daily lives.")

103. See Health Plays a Role in Older Adults' Vulnerability to Scams, Poll Suggests, Univ. of Mich. Inst. for Healthcare Pol'y & Innovation (Nov. 14, 2023), https://ihpi.umich.edu/news/health-plays-role-older-adults-vulnerability-scams-poll-suggests [https://perma.cc/SV5S-QVSR] ("'Our findings of a strong connection between scam vulnerability and health adds important new data to ongoing efforts to reduce the devastating toll of scams on older adults' finances and well-being,' said poll director Jeffrey Kullgreen . . . .").

reporting mechanisms.[104] These protocols can leverage evidence-based interventions to lighten the cognitive burden on users who are considering whether and how to report other users.[105] Another example of a "flow" is the steps that platforms impose on users who seek to change their UX to enhance their privacy. For example, today, in order to turn off Apple's AI capabilities—through which the company hones its AI technology by monitoring phone owners' activity on the applications in their phones—users must navigate through "Settings," identify "Apple Intelligence and Siri," select "Apps," and individually toggle off "Learn from this App" for each application.[106] A more intuitive "privacy flow" would allow users to disable Apple's AI monitoring of their devices in one toggle.

3. *Silencing Features*. — When users open their favorite social media platforms each day, they have the potential to interact with users from around the world. Yet those billions of profiles and pieces of content do not bombard their interfaces—the platform takes measures to moderate profile and content exposure. Similarly, platforms make design choices that impact users' own ability to regulate the profiles and content to which they are exposed. For example, "block" and "mute" design features on platforms permit users to reclaim and exercise autonomy over their UX.[107]

4. *Labels and Alerts*. — Social media companies can choose to create labels and alerts on various components of their UIs to draw users' attention to pertinent information. If users knew they were engaging with suspected scam content, they would be better equipped to avoid such schemes altogether. Due to the impact of disinformation and misinformation schemes on elites, social media companies have already taken strides to tackle political platform manipulation through platform design, including through the introduction of labels and "community

---

104. See Andrew Hutchinson, X Improves Content Reporting Flow, Making It Easier to Submit Rule-Breaking Content, Soc. Media Today (Sept. 24, 2023), https://www.socialmediatoday.com/news/x-improves-content-reporting-flow-making-easier-submit-rule-breaking/694568/ [https://perma.cc/P3CH-JYGN] ("[T]he new X reporting flow now gives you more specific violations to choose from when reporting a post. . . . Once your report is logged, you'll then be shown [a] screen highlighting possible actions you could take to limit any further harm.").

105. See Tom Muha, Opinion, Bye Bye, Blocking, Mich. Daily (Oct. 8, 2024), https://www.michigandaily.com/opinion/columns/bye-bye-blocking/ [https://perma.cc/LSP4-BVNC] (discussing statements by X's owner, who proclaimed a desire to eliminate the blocking feature from the social media platform).

106. Austin Williams, Apple's iOS 18.1 Brings AI Advancements: Privacy Tips You Need, Live Now Fox (Nov. 20, 2024), https://www.livenowfox.com/news/ios-18-1-ai-privacy [https://perma.cc/X4X9-42BP] (internal quotation marks omitted) (describing the steps users can take to ensure their "privacy remains intact" by disabling Apple's access to personal data).

107. Block, Mute, Restrict, Report—What's the Difference?, Instagram (Nov. 22, 2024), https://about.instagram.com/blog/tips-and-tricks/restrict-mute-block-report-guide (on file with the *Columbia Law Review*).

notes" on potentially misleading content.[108] Labels on content fall within the range of "publisher or speaker of third-party content" by the social media company that is protected by statutory immunity.[109] But labels are not limited to content. Platforms can design account labels, such as profile "verification" systems, that provide useful information to users. Research has also shown that scammers often send the same message to dozens or hundreds of targets at once; using signals like these, companies could detect suspected repeat offenders and create corresponding account labels.[110]

By monitoring suspected scam activity on their platforms and designing warning systems, social media companies can mitigate against platform-enabled deception. Warning messages that indicate whether a user has been previously reported for scams could put the community of social media users on notice of potential danger. Social media companies have this data; they routinely monitor online activity for groups suspected of dangerous activity.[111] Platforms can identify when individuals migrate communications off to third-party platforms and even identify AI-generated content.[112] Facebook notably created an image labeling system for AI-generated content in an effort to curb platform manipulation that could influence users' votes ahead of the 2024 U.S. presidential election; yet the platform offers no labeling system for similarly manufactured content that influences users to succumb to scammers.[113] Caution alerts on AI-generated content shared in direct messages could similarly put users on notice that they are dealing with scammers and reduce the psychological and reputational effects of this activity. For example, after

---

108. See Samantha Bradshaw, Shelby Grossman & Miles McCain, An Investigation of Social Media Labeling Decisions Preceding the 2020 U.S. Elections, PLOS ONE, Nov. 15, 2023, at 1, 7–9, 16, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0289683 [https://perma.cc/K2FJ-G4QY] (examining the impacts of labeling on Facebook and X and highlighting the need for platforms to permit Application Programming Interface access to allow researchers to further investigate platform dynamics).

109. Calise v. Meta Platforms, Inc., 103 F.4th 732, 740 (9th Cir. 2024) (internal quotation marks omitted) (quoting Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1101 (9th Cir. 2009)).

110. Noelle Toumey Reetz, Researchers Identify How Scammers Target Victims on Dating Apps, PHYS (Feb. 10, 2023), https://phys.org/news/2023-02-scammers-victims-dating-apps.html [https://perma.cc/AQC4-XJAS].

111. See Issie Lapowsky, Tech Companies Have the Tools to Confront White Supremacy, Wired (Aug. 14, 2017), https://www.wired.com/story/charlottesville-social-media-hate-speech-online/ (on file with the *Columbia Law Review*) (describing efforts by social media companies to combat white supremacist content and organizing on their platforms).

112. See Meta Identifies Networks Pushing Deceptive Content Likely Generated by AI, Reuters (May 29, 2024), https://www.reuters.com/technology/meta-identifies-networks-pushing-deceptive-content-likely-generated-by-ai-2024-05-29/ (on file with the *Columbia Law Review*) (finding that a Tel Aviv-based political marketing firm was behind a "covert influence operation[]" that weaponized generative AI).

113. Id.

the events of January 6th at the U.S. Capitol, Twitter (now X) deployed machine learning software to extricate violence-inducing content in record time.[114] Subsequent studies have verified the core proposition: Platforms could remove "habitual spreaders" of scam content in a heartbeat.[115]

5. *Pre-Bunking*. — "Pre-bunking," or "nudging," is a term that refers to the social media company practice of "inoculati[ng]" social media users from verified or suspected scam content.[116] This strategy "pre-emptively exposes people to tropes at the root of malicious [content], so they can better identify online falsehoods regardless of subject matter."[117] In turn, platform manipulators may be deterred from posting manipulative content; conversely, social media companies could point to this as conduct that satisfies their burdens of liability should negligence claims arise.

As another example, content algorithms curate the content that appears on users' timelines, but the *timing* of when that content is delivered is not a content decision. Rather, it is a platform design. Researchers have studied the optimal delivery of "pre-bunks" and have proposed new models for content delivery that can minimize users' likelihood of internalizing deceitful content, including messages.[118] By choosing when to show certain content, social media companies can design platforms that are less conducive to exploitation and deceit.

6. *Terms of Service Design*. — Beyond platforms themselves, several environmental and structural factors contribute to the industry of platform-enabled scams. Research has shown that consumers fell victim to more scams during the COVID-19 pandemic than in previous periods.[119]

---

114. Will Oremus, After Jan. 6, Twitter Banned 70,000 Right-Wing Accounts. Lies Plummeted., Wash. Post, https://www.washingtonpost.com/technology/2024/06/06/twitter-jan-6-deplatforming-misinfo-nature-study/ (on file with the *Columbia Law Review*) (last updated June 6, 2024).

115. Id.

116. Fred Lewsey, Social Media Experiment Reveals Potential to 'Inoculate' Millions of Users Against Misinformation, Univ. of Cambridge, https://www.cam.ac.uk/stories/inoculateexperiment [https://perma.cc/7AVA-4RCN] (last visited Jan. 24, 2025).

117. Id.; see also Tobias Rose-Stockwell, Facebook's Problems Can Be Solved With Design, Quartz (Apr. 30, 2018), https://qz.com/1264547/facebooks-problems-can-be-solved-with-design [https://perma.cc/YU6P-LM43] (describing four design choices for improving UX: "[g]iv[ing] [h]umanizing [p]rompts," "[p]icking out unhealthy content with better metrics," "[f]ilter[ing] unhealthy content by default," and "[g]iv[ing] users feed control").

118. See Yigit Ege Bayiz & Ufuk Topcu, Prebunking Design as a Defense Mechanism Against Misinformation Propagation on Social Networks 9 (Nov. 23, 2023) (unpublished manuscript), https://arxiv.org/pdf/2311.14200 [https://perma.cc/GF6B-HQUM] (finding an ideal algorithm for "optimally delivering prebunks").

119. See Monica T. Whitty, The Human Element of Online Consumer Scams Arising From the Coronavirus Pandemic, *in* Cybercrime in the Pandemic Digital Age and Beyond 57, 58 (Russel G. Smith, Rick Sarre, Lennon Yao-Chung Chang & Laurie Yiu-Chung Lau eds., 2023) ("[I]t is argued that the social and psychological conditions were, during the

Due to consumer psychology, consumers are falling prey to scams even when they agree to the terms and sign on dotted lines.[120] Social media companies that design their ToS to offer clear instructions for users can stifle platform manipulation by (re)alerting users of their rights and obligations. Importantly, ToS design does not refer to the content of the terms themselves—rather, it refers to how users interface with those terms.[121]

7. Account Verification Design. — Social media companies can also affect platform manipulation through their account verification policies. Through verification "badges" and other account badges, platforms introduce embellishments that can be exploited to the benefit of malicious actors.[122] Platforms engage in account verification design through the decisions they make concerning who can create an account[123] and how many accounts (and profiles) an individual or organization can create.[124] This is particularly relevant in the scam context, since scammers may share accounts, impersonate real accounts, and operate several accounts. One

---

height of the pandemic, very different to pre-COVID-19 times. It is most likely that these conditions account for some of the increase in the number of consumer scam[s] . . . .").

120. See Meirav Furth-Matzkin & Roseanna Sommers, Consumer Psychology and the Problem of Fine-Print Fraud, 72 Stan. L. Rev. 503, 510 (2020) ("[F]ine print may *disempower* consumers who read their contracts ex post . . . because consumers may become demoralized by contractual language and are likely to blame *themselves* for failing to read at the time of signing.").

121. See Designing the Terms and Conditions Page—Does It Really Matter? Yes It Does!, Encora (Sept. 25, 2019), https://insights.encora.com/insights/designing-the-terms-and-conditions-page [https://perma.cc/GF9S-JVC3] (describing ToS designs such as "[i]nformation grouping and structuring," summary sections with translations, "information popups," "icons and imagery," "[f]onts and spacing," Help sections, and FAQ formatting); Railslove, Terms of Service—An Opportunity in UX Design?, Medium (Nov. 15, 2018), https://medium.com/railslove/terms-of-service-an-opportunity-in-ux-design-2849e5fcea4e [https://perma.cc/4UT9-5QJD] (visualizing ToS designs that provide a poor user experience).

122. See, e.g., Craig Silverman and Bianca Fortis, Real Money, Fake Musicians: Inside a Million-Dollar Instagram Verification Scheme, ProPublica (Aug. 31, 2022), https://www.propublica.org/article/instagram-spotify-verified-fake-musicians [https://perma.cc/MWR9-32TU] ("[T]he operation transformed hundreds of clients into musical artists in an attempt to trick Meta . . . into verifying their accounts and hopefully paving the way to lucrative endorsements and a coveted social status.").

123. For example, platforms decide what age demographics can make an account. Many state legislatures have passed or are exploring age verification laws for social media companies. See Jenna Zhang, Lindsey Tonsager, Diana Lee, Madeline Salinas & Priya Leeds, State, Federal, and Global Developments in Children's Privacy, Q1 2023, Covington (Apr. 2, 2023), https://www.insideprivacy.com/childrens-privacy/state-federal-and-global-developments-in-childrens-privacy-q1-2023/ [https://perma.cc/9QFL-MUVA] (describing Utah's law "requiring social media companies to verify the age of all users to determine which are under eighteen").

124. See, e.g., FE Tech Desk, Facebook Testing Feature to Allow Users to Have Up to Five Profiles, Fin. Express (July 15, 2022), https://www.financialexpress.com/life/technology-facebook-additional-profiles-feature-test-meta-platforms-2595469 [https://perma.cc/3DPU-SQPF].

platform that allows users to video chat with strangers, Omegle, was sued for negligent design choices that matched an eleven-year-old girl with a thirty-year-old man who would come to sexually abuse her for years.[125] The platform's decision not to verify accounts before making connections between adults and minors is an example of a design choice that works to the advantage of malicious actors. Similarly, Grindr, a dating application, has faced lawsuits over its negligent design of an age verification process that promoted grooming.[126] But courts thus far have held that, for claims related to account verification processes, courts cannot treat social media companies like "publishers" of account data, which invokes immunity for platforms and forecloses liability.[127]

C. *The Platform Manipulation Economy*

Platforms often generate revenue from advertising and selling user data, which incentivizes them to respond to user expectations insofar as those responses lead users to spend more time on, and engage with, their platforms.[128] Scams and other platform manipulation corollaries disrupt the notion that companies design platforms to "match[] users' expectations, [so that] users will spend more time on the site and advertising revenue will increase."[129] This logic presumes that companies are able to accurately meet user expectations, and it neglects the core misalignment that scammers and malicious actors capitalize on: Sometimes companies' perceptions of users' expectations are distorted (and users' self-perceptions can be distorted). The platform economy, as robust and multidimensional as it has become,[130] continues to thrive on platform designers' limited constructions of user expectations. Even while designing UXs, social media companies tend to experiment with large groups,[131] which can neglect the experiences of minorities and other marginalized communities online.

---

125. See A.M. v. Omegle.com, LLC, 614 F. Supp. 3d 814, 817 (D. Or. 2022).

126. See, e.g., Doe v. Grindr Inc., 709 F. Supp. 3d. 1047, 1050–51 (C.D. Cal. 2023); Nazgole Hashemi & Tannaz H. Hashemi, Don't Let Them Fool Ya: An Examination of Regulation Crowdfunding as a Framework for Federal Protection Against Online Dating Risks, 53 U.S.F. L. Rev. 421, 423 (2019) ("Negligence cases against online dating platforms are subject to dismissal because the law currently imposes no duty on them to conduct criminal background checks or otherwise take steps to ensure the safety of users.").

127. See Hashemi & Hashemi, supra note 126, at 422–23.

128. Engagement can include everything from opening the platform's webpage to clicking on links to exchanging messages with a scammer. See Kate Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, 131 Harv. L. Rev. 1598, 1627 (2018) [hereinafter Klonick, New Governors] (describing the engagement-based platform economy).

129. Id.

130. See infra notes 187, 189 and accompanying text.

131. See Konstantino, supra note 99, at 189 n.56 ("Traditionally, companies segment users into two groups at random and show each group one of two versions of the app. Recently, testing has gotten more complex to account for confounding variables . . . .").

Platform manipulators' interests converge[132] with platforms' interests in a way that leads to devastating effects for victims of online scams, disinformation and misinformation campaigns, and other kinds of platform-based deception. These actors need users to spend more time interacting with them on platforms in order to develop stronger deception-based relationships.[133] Thus, social media companies can profit immensely from platform manipulation: When users spend more *time* on the platform, the company can "sell" those numbers to advertisers in order to generate revenue.[134] These companies may also be able to profit politically from remaining silent or refusing to raise the alarm on issues that impact their reputation and standing with stakeholders, including lawmakers.[135] Moreover, the status quo laissez-faire approach to social media regulation "invites the worst abuses by the state."[136] In addition to earning revenue from social media scams through metrics sold to advertisers, social media companies save money by not addressing platform manipulation in the short-term;[137] tackling this issue requires hard-to-find, multifaceted expertise in UX design and scams, disinformation, and other areas.[138]

Social media companies play a crucial and foundational role in the platform economy. Due to the global nature of these schemes and the ability to hide identities online, it is extremely difficult to go after

---

132. See generally Derrick A. Bell, Jr., Comment, *Brown v. Board of Education* and the Interest-Convergence Dilemma, 93 Harv. L. Rev. 518 (1980) ("[T]his principle of 'interest convergence' provides: The interest of blacks in achieving racial equality will be accommodated only when it converges with the interests of whites").

133. See Pistone & Knowles, supra note 4 (describing "pig-butchering" scams predicated on the duration of time for their efficacy).

134. See Leetaru, supra note 23 ("Facebook is in reality renting access to data. Its sole value proposition to developers is access to its two billion users.").

135. See Michael L. Rustad & Thomas H. Koenig, Rebooting Cybertort Law, 80 Wash. L. Rev. 335, 345 (2005) ("Corporate stakeholders use their lobbying influence to expand their online rights and to avoid liability."); David Greene, In These Five Social Media Speech Cases, Supreme Court Set Foundational Rules for the Future, Elec. Frontiers Found. (Aug. 14, 2024), https://www.eff.org/deeplinks/2024/08/through-line-suprme-courts-social-media-cases-same-first-amendment-rules-apply [https://perma.cc/4ZUH-DT2P] (describing several high-profile Supreme Court cases involving social media companies, including cases concerning the interdependence between lawmakers and social media companies).

136. See Kyle Langvardt, Regulating Online Content Moderation, 106 Geo. L.J. 1353, 1386 (2018) ("[The social media system] mediates a dominant and growing share of all online communication, and its private owners are few enough in number to operate as convenient 'choke points' under pressure.").

137. See supra note 23 and accompanying text (discussing the long-term value proposition for platforms that combat scams).

138. See Rob Rashotte, Why Closing the Cyber Skills Gap Requires a Collaborative Approach, World Econ. F. (July 23, 2024), https://www.weforum.org/stories/2024/07/why-closing-the-cyber-skills-gap-requires-a-collaborative-approach/ (on file with the *Columbia Law Review*) (describing the global cybersecurity labor shortage).

deceptive actors themselves.[139] The lack of adequate remedies against these primary violators leaves social media companies at the leading edge of both harm perpetration and potential recourse for victims. In the next Part, this Note analyzes the shortcomings of the existing liability frameworks available to the individuals and groups on the other end of platform manipulation.

II. THE SHORTCOMINGS OF EXISTING LIABILITY FRAMEWORKS

Platform manipulation takes several forms and thrives on many aspects of platforms, including platform design. Often, platform manipulators leverage platform design elements to implement their schemes. Considering this development—made possible by the innovation of network-effects and social media–platform technologies in the past three decades—§ 230 (discussed in section II.A), consumer law (section II.B), and voluntary self-regulation (section II.C) are woefully maladapted to confront platform manipulation.

A. *Platform Design as Content-Agnostic Corporate Conduct: The § 230 Immunity Myth*

The centerpiece of social media law, § 230 of the Communications Decency Act of 1996,[140] persists as the bulwark against social media–company liability for the harms their platforms cause to users by way of content moderation decisions.[141] This does not mean, however, that social media companies cannot be held liable for other harms caused to their

---

139. See Teele et al., supra note 61.

140. Section 230 broadly provides "internet service providers" (i.e., social media companies) with broad immunity over their decisions to keep, promote, downgrade, and remove content, as well as their decisions to suspend or ban users.

> 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. . . .
>
> No provider or user of an interactive computer service shall be held liable on account of—
>
> > (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be . . . objectionable, whether or not such material is constitutionally protected . . . .

47 U.S.C. § 230(c) (2018).

141. In the early days of website hosting, two New York cases played exceedingly influential roles in shaping the contours of "social media law." For an overview of "social media law," see generally Social Media Law Bulletin, Norton Rose Fulbright LLP, https://www.socialmedialawbulletin.com/glossary-of-us-laws/ [https://perma.cc/S5BD-YHVT] (last visited Jan. 28, 2025); see also Cubby, Inc. v. CompuServe Inc., 776 F. Supp. 135, 141 (S.D.N.Y. 1991) (holding that an online messaging board was not liable for content it was not aware of); Stratton Oakmont, Inc. v. Prodigy Servs. Co., 1995 WL 323710, at *5 (N.Y. Sup. Ct. May 23, 1995) (holding that a separate online messaging board was liable for content on its site because it had attempted to moderate *some* posts).

consumers. Countless law enforcers and private plaintiffs have sued social media companies in relation to platform-based deception.[142] Unfortunately, due to misunderstandings of § 230 and applications of the First Amendment to technology platforms, some scholars have continued to portray that social media platforms are entirely immune for their non-content-related decisions—including their platform design choices.

At the same time, courts are increasingly recognizing that § 230's safe harbor does not shield all platform conduct from liability. The Third Circuit recently held that "TikTok's recommendations via its ["For You Page" timeline] algorithm . . . [was] TikTok's own expressive activity," subject to liability under § 230.[143] The case surrounded a ten-year-old girl, Nylah Anderson, who died after participating in a "Blackout Challenge" algorithmically advertised to her by TikTok.[144] The construction of algorithms that recommend content invokes numerous platform design levers, namely those that permit users to play a role in "boosting" or "suppressing" content in the algorithm.[145] Ultimately, it is extremely difficult for outsiders to determine whether social media companies are taking content-neutral or content-responsive decisions when developing or editing their algorithms, as has been the case with accusations of platform censorship for politically divisive topics.[146]

In effect, § 230 is quite vague; the law provides no definitions for "good faith" content moderation or "objectionable" material, despite mentioning the former and policing the latter.[147] Critics in both the Democratic and Republican parties have unsuccessfully sought to both expand and curtail the reach of the statute, while simultaneously

---

142. See Lemmon v. Snap, Inc., 995 F.3d 1085, 1092 (9th Cir. 2021) ("[T]he Parents' amended complaint does not seek to hold Snap liable for its conduct as a publisher or speaker. Their negligent design lawsuit treats Snap as a products manufacturer . . . negligently designing a product (Snapchat) with a defect (the interplay between Snapchat's reward system and the Speed Filter)."); Doe v. Grindr Inc., 709 F. Supp. 3d 1047, 1050 (C.D. Cal. 2023) ("Doe brings this lawsuit against Grindr for child sex trafficking and a defective product, asserting claims of strict product liability, negligence, negligent misrepresentation, and violation of the Trafficking Victims Protection Reauthorization Act . . . .").

143. Anderson v. TikTok, Inc., 116 F.4th 180, 184 (3d Cir. 2024).

144. Id. at 181.

145. See Danielle Draper, Demystifying Social Media Algorithms, Bipartisan Pol'y Ctr. (Aug. 2, 2023), https://bipartisanpolicy.org/blog/demystifying-social-media-algorithms (on file with the *Columbia Law Review*) (describing design levers such as the use of "viewing history, likes, shares, comments, accounts followed, demographics, geographic location, preferences, and search history" to control the kind of content displayed to users).

146. See Priyanka Shankar, Pranav Dixit & Usaid Siddiqui, Are Social Media Giants Censoring Pro-Palestine Voices Amid Israel's War?, Al Jazeera (Oct. 24, 2023), https://www.aljazeera.com/features/2023/10/24/shadowbanning-are-social-media-giants-censoring-pro-palestine-voices [https://perma.cc/9FQN-57N9] (describing a "bug" that led Meta to decrease exposure of social media posts that included mentions of Palestine).

147. Edward Lee, Moderating Content Moderation: A Framework for Nonpartisanship in Online Governance, 70 Am. U. L. Rev. 913, 925 (2021) (internal quotation marks omitted).

expressing a desire for greater accountability for social media companies.[148] Above all, § 230 does not prevent public and private parties from ascribing liability to social media companies for their platform design choices. The statute's clear aim at "action voluntarily taken in good faith to restrict access to or availability of material,"[149] or protection for content-based restrictions, is wholly detached from platforms' *design* decisions.

Platform design choices that enable platform manipulation fall outside § 230's purview for one principal reason.[150] The conduct at issue in such cases[151]—platform design choices—does *not* serve the purpose of restricting the availability of objectionable material. Rather, these platform design choices are made in order to connect users to one another, retain user attention to the platform, and contribute to the overall UX, which are all not forms of "content." Courts have stated as much when the platform design choice to provide verification badges to hijacked YouTube channels fell outside the scope of § 230.[152] Nonetheless, scholars have continued to

148. In 2021, Democratic lawmakers introduced the "Health Misinformation Act of 2021," seeking to hold companies liable when they allow "health misinformation" to proliferate on their platforms. S. 2448, 117th Cong. (2021). Republican bills include the "Online Freedom and Viewpoint Discrimination Act," which would modify § 230 to limit protections for platforms. S. 4534, 116th Cong. (2020). For a more comprehensive list of all § 230-related bills, see All the Bills on Section 230, Civic Genius (Feb. 9, 2022), https://www.ourcivicgenius.org/learn/all-the-bills-on-section-230/ [https://perma.cc/MSA6-2KEV]; Chris Riley & David Morar, Legislative Efforts and Policy Frameworks Within the Section 230 Debate, Brookings Inst. (Sept. 21, 2021), https://www.brookings.edu/articles/legislative-efforts-and-policy-frameworks-within-the-section-230-debate/ [https://perma.cc/AHZ4-HBZK]. For a critique of Democratic and Republican approaches, see Tim Wu, Liberals and Conservatives Are Both Totally Wrong About Platform Immunity, Medium (Dec. 3, 2020), https://superwuster.medium.com/liberals-and-conservatives-are-both-totally-wrong-about-section-230-11faacc4b117 [https://perma.cc/N5CM-RQ8K] (describing the challenges associated with an all-or-nothing approach to Section 230 reform).

149. 47 U.S.C. § 230(c)(2)(A) (2018).

150. See, e.g., Danielle Keats Citron, Section 230's Challenge to Civil Rights and Civil Liberties, Knight First Amend. Inst. (Apr. 6, 2018), https://knightcolumbia.org/content/section-230s-challenge-civil-rights-and-civil-liberties [https://perma.cc/KT46-U3FG] ("Platforms disadvantage the vulnerable not just through their encouragement of cyber mobs and individual abusers but also through their design choices. . . . Section 230 should not be read to immunize platforms from liability related to user interface or design."). Such critiques of § 230's disassociation from design choices have centered around discrimination, harassment, and illegal behaviors facilitated by platforms, as opposed to consumer scams and other platform-based manipulation. See id. ("When code enables invidious discrimination, law should be allowed to intervene."); Olivier Sylvain, Discriminatory Designs on User Data, Knight First Amend. Inst. (Apr. 1, 2018), https://knightcolumbia.org/content/discriminatory-designs-user-data [https://perma.cc/AT2Z-PK2U] (arguing that "courts should account for the specific ways in which intermediaries' designs do or do not enable or cause harm to the predictable targets of discrimination and harassment").

151. See infra notes 236–243 and accompanying text.

152. The court could not rule on this issue because the plaintiffs had not pleaded this argument. See Wozniak v. YouTube, LLC, 319 Cal. Rptr. 3d 597, 603 (Cal. Ct. App. 2024) ("[W]e also conclude that one of plaintiffs' claims—that defendants created their own

ascribe a broader meaning to § 230 than exists within the text of the statute.[153]

The legislative history of § 230 demonstrates that the law at its inception was not designed to apply to platform design choices. Passed by a margin of 420-4, § 230 was intended for two purposes: to "encourage the unfettered and unregulated development of free speech on the Internet" and to empower platforms to police their content and address child safety on the internet.[154] Importantly, § 230 was *not* intended to apply to the architecture of platforms, their context-agnostic presentation of content, their capacity to detect malicious actors, their responsibilities in relation to the information that they adduce from their platforms, or anything of the like. Representative Christopher Cox, co-author of § 230 alongside Representative Ron Wyden, wrote in an amicus brief for the 2022 case of *NetChoice, LLC v. Florida* that "the plain meaning of the words in Section 230 is exactly what Congress intended."[155] It was intended to "establish[] clear rules of liability tailored to the essential characteristics of the Internet in order to expand opportunities for users to create and publish their own content."[156] According to Representative Cox, the law was intended to apply to platforms acting "as arbiters of content moderation" that could help cultivate a "broad range of interests, each with its own community standards."[157]

Section 230 was also written with a particular bent on preserving the safety of children on the internet. Considering copious evidence illustrating the widespread nature of platform-enabled scams, which disproportionately target elderly individuals, it is difficult to imagine the architects of § 230 would have meant to remove liability for when reporting flows contribute to elder abuse scams. Unfortunately, in limited instances, scholars have adopted an atextual interpretation of § 230 to foreclose liability over platform design.[158] Meanwhile, § 230 may continue

---

content and materially contributed to the unlawfulness of the scam by providing verification badges to hijacked YouTube channels—includes allegations which potentially could fall outside the scope of section 230 immunity.").

153. See Hashemi & Hashemi, supra note 125, at 422 ("Holding dating platforms liable for third-party misconduct is virtually impossible at this time, although they are responsible for facilitating connections.").

154. Section 230: Legislative History, Elec. Frontiers Found., https://www.eff.org/issues/cda230/legislative-history [https://perma.cc/EN58-TGT2] (last visited Jan. 25, 2025) (internal quotation marks omitted) (quoting Bratzel v. Smith, 333 F.3d 1018, 1033 (9th Cir. 2003)).

155. Brief of Former U.S. Representative Christopher Cox, Co-Author of Section 230, as Amicus Curiae in Support of Conditional Cross-Petitioners at 2, NetChoice, LLC v. Moody, 114 S. Ct. 69 (mem.) (2023) (No. 22–393), 2022 WL 17338954, cert denied.

156. Id. at 3.

157. Id.

158. See Allison M. Clay, Comment, Blissful Unaccountability: The Nonregulation of Precarious Network Marketing Schemes on Social Media, 47 Del. J. Corp. L. 595, 605 (2023) (claiming that, because of Section 230, "regardless of the role of social networking platforms

to cost the public access to public spheres.[159] Though platforms' scam monitoring activities fall outside the bounds of their statutory immunity, to date, no plaintiffs have advanced a theory of liability that argues that platforms owe users a responsibility to inform them when they use the platform to maintain a relationship with an individual previously reported for fraudulent or scam activity.

B. *Platform Design as a Duty: U.S. Consumer Law's Neglect of User Rights*

Broadly speaking, U.S. consumer law fails to protect users' rights, including their rights in private litigation involving platform manipulation.[160] Today, there is no statutorily enshrined right of action available to plaintiffs at the state or federal level that appreciates a consumer's right to reasonable, safe, or protective platform designs.[161] Rather, a patchwork of laws governs platform manipulation. The main sources of law are the Federal Trade Commission Act (FTCA), the Consumer Review Fairness Act, and cyber exploitation-focused laws like the Children's Online Privacy Protection Act (COPPA). The primary enforcers are the FTC and Consumer Financial Protection Bureau (CFPB). By and large, current enforcement efforts have fallen short in the task of ascertaining platform liability for platform manipulation.

Through its authorities under § 5 of the FTCA, the FTC is responsible for pursuing relief for consumer-victims of "injurious conduct."[162] In its

---

in facilitating MLMs and pyramid schemes, they cannot be held accountable under the law for the harm that these schemes cause their users").

159. See David Pozen, Intermediary Immunity and Discriminatory Designs, Knight First Amend. Inst. (Apr. 6, 2018), https://knightcolumbia.org/content/intermediary-immunity-and-discriminatory-designs [https://perma.cc/97HZ-VLC9] ("[Section 230] has arguably shaped the development of the public sphere in problematic ways—subsidizing digital platforms over analog ones, rewarding reliance on user-generated rather than employee-generated content, and allowing website operators to avoid internalizing many of the social costs of the materials they disseminate.").

160. See Roger Allan Ford, Data Scams, 57 Hous. L. Rev. 111, 142 (2019) ("Although many scams violate the law, there are enough that are legal, or that are not clearly illegal, that existing law is not a reliable solution to the problem of targeted scams.").

161. See infra Part III.

162. Katherine Waitz, Comment, A Shift in the Tides? The Welcomed Proposal of Harshened FTC Guidelines for Social Media Reviews and Advertising, 51 S.U. L. Rev. 129, 132 (2023). The FTCA governs platform manipulation that involves commercial transactions. Under the FTCA, the FTC must act by:

> (a) preventing unfair methods of competition and unfair or deceptive acts or practices in or affecting commerce; (b) seeking monetary redress or other relief for injurious conduct to consumers; (c) prescribing rules defining acts or practices that are unfair or deceptive, and establishing requirements designed to prevent such acts or practices; and (d) gathering and compiling information and conducting investigations relating to such practices, organizations, businesses, and management of entities engaged in commerce.

Id; see also 15 U.S.C. § 45 (2018).

focus on the social media space, the FTC has largely targeted social media influencers, advertisers, and companies that engage in deceptive marketing practices,[163] though lay consumers are the prototypical victims of platform manipulation.[164] Historically the FTC viewed social media harms through the lens of privacy and security,[165] which often accompany and may be ancillary to the financial, reputational, and psychological harms caused by deceptive online conduct.[166]

Contemporary legal framing of platform manipulation nascently posits platform manipulation and platform design as "deceptive acts" and "unfair methods" under the FTCA[167] and similar state laws, pursuant to the FTC's authority to seek relief for injuries arising from platform manipulation and platform design insofar as users are social media consumers.[168] The CFPB is another federal agency with a similar mission of ensuring "that markets for consumer financial products and services are fair, transparent, and competitive,"[169] though its ability to respond to deceptive practices has been weakened; notably, the agency has previously taken action to subvert efforts to undermine student loan scams operating on social media.[170]

Through the Consumer Review Fairness Act, which was passed in 2016,[171] Congress has taken action to curb platform designs that exclude negative product reviews, and the FTC has used its enforcement power to

---

163. See, e.g., Press Release, FTC, Fashion Nova Will Pay $4.2 Million as Part of Settlement of FTC Allegations It Blocked Negative Reviews of Products (Jan. 25, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/01/fashion-nova-will-pay-42-million-part-settlement-ftc-allegations-it-blocked-negative-reviews [https://perma.cc/U4HS-TPDC].

164. See Julie Brill, Privacy & Consumer Protection in Social Media, 90 N.C. L. Rev. 1295, 1296 (2012) (discussing consumer protection issues caused by the way social media has "changed the way companies do business and the way they interact with consumers").

165. See id. at 1299 ("We continue to monitor the social media space for practices that impact the privacy and security of the personal information about consumers.").

166. See supra section I.A.

167. See 15 U.S.C. § 45.

168. For example, in July 2024, the U.S. Department of Justice filed a complaint against global software company Adobe, Inc. for, among other things, obscuring the terms of its "'Annual, Paid Monthly' subscription plan" using an "onerous and complicated cancellation process" and "optional textboxes and hyperlinks, providing disclosures that are designed to go unnoticed and that most consumers never see." Complaint for Permanent Injunction, Monetary Judgment, Civil Penalty, and Other Relief at 2, United States v. Adobe Inc., No. 5:24-cv-03630-BLF (N.D. Cal. filed July 23, 2024), 2024 WL 3680811 (internal quotation marks omitted).

169. 12 U.S.C. § 5511(a) (2018); see also id. § 5491(a) (establishing the CFPB to "regulate the offering and provision of consumer financial products or services under the Federal comsumer financial laws").

170. See Creola Johnson, Relief for Student Loan Borrowers Victimized by "Relief" Companies Masquerading as Legitimate Help, 11 U.C. Irvine L. Rev. 105, 144–51 (2020) (explaining how CFPB leadership under acting director Mick Mulvaney "implemented several changes deemed harmful to student loan borrowers").

171. Consumer Review Fairness Act of 2016, Pub. L. No. 114–258, 130 Stat. 1355 (codified at 15 U.S.C. § 46(b)).

curb similar conduct involving fake reviews.[172] The FTC has also taken action to limit the selling of fraudulent or deceptive products, but has not yet posited consumer time spend as a transaction that elicits liability for platform design.[173] Importantly, FTC and state laws on deceptive advertising fail to conceptually account for a robust definition of platform manipulation because they are generally limited to conduct "affecting commerce."[174] While platform manipulation victims are often deceived about the purpose for engaging in commercial transactions, in romance and other scams, victims transfer money directly into scammers' bank accounts. In addition, FTC enforcement is hampered by the difficulties associated with identifying perpetrators due to the frequently transnational, subtle, and hard-to-detect nature of platform manipulation.[175]

As understood by legal actors and consumers, consumer protection law cannot regulate the "false speech of private citizens in non-commercial settings."[176] While scammers and other platform manipulators engage in "false speech," platform manipulation, platform design, and platforms themselves are not yet widely understood as commercial settings.[177] Unfortunately, the conditions for this reality are well-documented—both "the United States and other jurisdictions have not undertaken systemic reviews of their consumer protection regimes to ensure they are fit for the challenges . . . in online markets."[178] Efforts to revamp consumer protec-

---

172. See Andrea M. Matwyshyn & Miranda Mowbray, Fake, 43 Cardozo L. Rev. 643, 659 (2021) (describing the application of the Consumer Review Fairness Act to delicately navigate the "complex" legal questions behind "intent and quantification of harm" in the fake reviews context).

173. See Nicole Dunn, Note, A Dupe or Just Duped? An Analysis of the History and Policy Behind Counterfeit Cosmetics and Social Media's Role in Perpetuating Its Sales, 20 J. Health & Biomedical L. 92, 100–04 (2024) (describing the FTC's authority to police fraudulent business practices that transpire online).

174. See, e.g., 15 U.S.C. § 45(a)(1) ("Unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are hereby declared unlawful."); Cal. Bus. & Prof. Code § 17508 (2024) ("It shall be unlawful for any person doing business in California and advertising to consumers in California to make any false or misleading *advertising* claim . . . ." (emphasis added)); Iowa Code § 714.16(2)(a) (2025) (prohibiting deception "in connection with the lease, sale, or advertisement of any merchandise or the solicitation of contributions for charitable purposes"); N.Y. Penal. Law § 190.20 (McKinney 2025) ("A person is guilty of false advertising when, with intent to promote the sale or to increase the consumption of property or services, he makes or causes to be made a false or misleading statement in any advertisement . . . .").

175. See Ford, supra note 160, at 168–72 ("A key challenge in implementing law-enforcement tools, then, will be overcoming that lack of technical expertise.").

176. Ira Rubinstein & Tomer Kenneth, Taming Online Public Health Misinformation, 60 Harv. J. on Legis. 219, 245 (2023).

177. See supra section I.C. (offering the platform economy as a commercial setting); infra Part III (introducing the Platform Design Negligence paradigm).

178. Amelia Fletcher et al., Consumer Protection for Online Markets and Large Digital Platforms, 40 Yale J. on Regul. 875, 879 (2023) ("The failure to update consumer-protection law is concerning in part because we rely on it to advance a broad range of interests in addition [to] the purely economic interests of market participants.").

tion laws for the platform economy have been unsuccessful,[179] potentially due to outmoded conceptions of contemporary scams and frauds.[180]

Identity theft protection laws, such as the Identity Theft and Assumption Deterrence Act of 1998,[181] are hard to apply given the difficulties with identifying perpetrators who are often located outside the U.S. While these laws are helpful in cases of celebrity impersonations, most consumer scams do not involve impersonation of the victim. For celebrity impersonator scams, celebrities are neither necessarily incentivized nor able to sue on the victims' behalf. In cases involving lay individuals, the FTC has pursued enforcement action against companies like Match.com for presenting fake profiles to entice users as a form of UX design for user recruitment and retention.[182] The Match.com case, which has been pending for over five years, offers one opportunity for the U.S. District Court for the Northern District of Texas to recognize platform liability for platform manipulation.[183]

Cyber exploitation—including both instances when intimate partners share sexually explicit images and other content without consent from the individuals depicted in the content and AI-generated sexually explicit content of real individuals—is an area in which the FTC and state enforcers have tried to act.[184] Similarly, law enforcers have focused on the impact of social media on children, strengthening enforcement of laws like COPPA.[185]

---

179. See David Adam Friedman, Reinventing Consumer Protection, 57 DePaul L. Rev. 45, 46 (2007) ("Policymakers can neither transform the entire consumer protection system overnight nor allocate more resources to the problem.").

180. See Friedman, Impostor Scams, supra note 44, at 58 ("As technology evolves, new, corporate-driven products and services become increasingly difficult to understand. As stand-alone swindlers develop new schemes, regulators will constantly fail to think ahead of the perpetrators.")

181. 18 U.S.C. § 1028 (2018).

182. Press Release, FTC, FTC Sues Owner of Online Dating Service Match.com for Using Fake Love Interest Ads to Trick Consumers Into Paying for a Match.com Subscription (Sept. 25, 2019), https://www.ftc.gov/news-events/press-releases/2019/09/ftc-sues-owner-online-dating-service-matchcom-using-fake-love [https://perma.cc/BLV9-W7VM].

183. Match Group, Inc., FTC, https://www.ftc.gov/legal-library/browse/cases-proceedings/172-3013-match-group-inc [https://perma.cc/9552-ARM2] (last updated Sept. 25, 2019).

184. See Nonconsensual Distribution of Intimate Images: What to Know, FTC (Nov. 2024), https://consumer.ftc.gov/articles/nonconsensual-distribution-intimate-images-what-know [https://perma.cc/DF6A-ZSDZ] (sharing a resource with information about state laws and requesting that victims report incidents of nonconsensual image sharing to the FTC).

185. See Brill, supra note 164, at 1299–304 ("The implications of COPPA in the social media context are significant. Social media operators subject to COPPA must obtain parental consent prior to the collection, use, or disclosure of information about children."); Cole F. Watson, Protecting Children in the Frontier of Surveillance Capitalism, 27 Rich. J.L. & Tech. at 1, 5 (2021) (arguing for COPPA reforms that are responsive to the "unprecedented acceleration of the digital frontier").

C. *Platform Design as Governance: Deconstructing Voluntary Self-Governance*

Many factors play into the failure of the law to meaningfully grapple with social media companies' complicity in platform manipulation. These dynamics are reproduced by the logic of self-governance that can muddy the clear lines between content-based and platform-design decisions

Social media companies have evolved into sophisticated entities capable of operating full-scale marketplaces,[186] even enabling organized criminal organizations to launder money[187] and creators to monetize adult content.[188] The internet behavior of social media users has also changed.[189] New wholesale models for social, economic, and cultural ordering, also known as the "platform economy," provide platforms with endless possibilities for framing their own social obligations.[190]

Against that backdrop, "platform governance" has emerged as a prevailing paradigm for conceiving of the relationship between social media companies and the actors that abuse their platforms.[191] It "refers to the policy, technical, and design decisions impacting a global network of internet users."[192] It portrays social media companies as counterparts to

---

186. See How Marketplace Works, Facebook, https://www.facebook.com/help/1889067784738765 [https://perma.cc/SGV4-TV83] (last visited Jan. 25, 2025) (describing how Meta users can post "listings" through the "Marketplace" platform).

187. See Rohena Rajbhandari, Note, (Ven)mo Money, (Ven)mo Problems? How Money Laundering Permeates Peer-to-Peer Payment Platforms, 63 B.C. L. Rev. 669, 671 (2022) ("Despite the United States' robust anti-money laundering laws, concerns regarding money laundering still permeate the P2P market, as existing state and federal laws lack clarity and do not fully address emerging concerns.").

188. "Creators," often referred to as "influencers," are defined as individuals who generate content that they can monetize in the platform economy. Bernhard Rieder, Erika Borra, Òscar Coromina & Ariadna Matamoros-Fernández, Making a Living in the Creator Economy: A Large-Scale Study of Linking on YouTube, 9 Soc. Media + Soc'y, Apr.–June 2023, at 1, 1.

189. See Mary Aiken, The Cyber Effect 18 (2016) (applying the discipline of cyberpsychology to shine a light on how "behavior mutates in cyberspace").

190. Lucy Colback, The Rise of the Platform Economy, Fin. Times (Mar. 13, 2023), https://www.ft.com/content/e5f5e5b9-3aec-439a-b917-7267a08d320f [https://perma.cc/6CHL-L4FK].

191. Platform governance was the subject of *The New Governors*, a 2018 *Harvard Law Review* article that provided a conception of how social media platforms adapt and operate in a rapidly changing internet ecosystem. Klonick, New Governors, supra note 128, at 1602, 1662. While the term has been applied to non-social media platform-based businesses, this Note uses "platform governance" to specifically refer to social media platforms. See Susan Etlinger, The Next Wave of Platform Governance, Ctr. for Int'l Governance Innovation (May 14, 2021), https://www.cigionline.org/articles/next-wave-platform-governance/ [https://perma.cc/7X3R-ZTNW] ("Because each platform type—advertising, cloud, industrial, product, lean—has a distinct set of characteristics, products, services, ways of making money and relative risk, each carries a distinct set of governance implications as well.").

192. Introducing an ISP-WIII Essay Series Exploring the Terms and Concepts that Constitute Platform Governance., Yale L. Sch. Info. Soc'y Project, https://law.yale.edu/isp/publications/platform-governance-terminologies [https://perma.cc/V9VF-JRZT]

government agencies that borrow principles from administrative law and further "democratic culture."[193] The paradigm propagates an assumption about a collective good that obscures the nature of the individualized relationship between platforms and users.

Platform governance is a popular—if not "existential"[194]—container for legal scholars to espouse interpretations of and proposals relating to the power of platforms.[195] Though the term "governance" accompanies conventional narratives of platform capitalism that further prevailing neoliberal economic accounts of platforms,[196] legal scholarship in this area

---

[hereinafter Terms and Concepts that Constitute Platform Governance] (last visited Jan. 26, 2025).

193. Klonick, New Governors, supra note 128, at 1663.

194. See Charilaos Papaevangelou, The Existential Stakes of Platform Governance: A Critical Literature Review 4 (July 1, 2021) (unpublished manuscript), https://doi.org/10.12688/openreseurope.13358.2 [https://perma.cc/FV39-TB7W] (using the paper "to surface an existential risk that lies with the way that current scholarship approaches platform regulation and governance: that of conflating the internet with large social media platforms").

195. The concept of platforms as "governors" was coined by Kate Klonick in her seminal 2018 article *The New Governors: The People, Rules, and Processes Governing Online Speech*. Klonick, New Governors, supra note 128 at 1603. The article was the first of its kind to provide an in-depth legal analysis of social media companies, which Klonick achieved through original interviews with current and former employees of X and Meta (formerly Facebook), as well as "internal documents" she was directly provided by Meta. Id. at 1602. Such access may have contributed to the article's explosive success. Cf. Brenda Dvoskin, The Illusion of Inclusion: The False Promise of the New Governance Project for Content Moderation, 93 Fordham L. Rev. 1315, 1325 (2025) (calling *The New Governors* an "influential piece" that "was the beginning of an explosion of legal scholarship in the content moderation field"). The article also advances a generous claim that "platforms play no significant role—yet—in determining whether content is true or false." Klonick, New Governors, supra note 128, at 1660 (footnote omitted). While platforms may not play an explicit role in determining whether content is true or false, platforms do play a significant and explicit role in determining what content to flag as "misleading content." See, e.g., Community Notes: A Collaborative Way to Add Helpful Context to Posts and Keep People Better Informed, X, https://communitynotes.x.com/guide/en/about/introduction [https://perma.cc/G3XH-4R9D] (last visited Jan. 25, 2025) (explaining that while community users are the ones flagging content as misleading, X maintains control over which of those flags appears to other users). In *The New Governors*, Klonick provided a curated look into how social media companies make decisions about the environment on their platforms. See Klonick, New Governors, supra note 128, at 1669 ("Through interviews with former platform architects and archived materials, this Article argued that platforms moderate content partly because of American free speech norms and corporate responsibility, but most importantly, because of the economic necessity of creating an environment that reflects the expectations of their users."). She argued their approach was informed by well-intentioned lawyers who crafted platforms' content moderation policies in reliance on the First Amendment and free speech principles. Id. at 1660.

196. See, e.g., Frank Pasquale, Two Narratives of Platform Capitalism, 35 Yale. L. & Pol'y Rev. 309, 311–15 (2016). One example of a conventional narrative is that "[l]arge digital platforms have gained massive market share because of the quality of their service," whereas the counternarrative proposed by Pasquale says, "[l]arge digital platforms have gained massive market share because of luck, first-mover advantage, network effects,

has generally embraced the "governance" framework for conceiving of *how* platforms' decisions are made.[197] Thus, the platform governance framework sits directly at odds with the tort framework provided by the Platform Design Negligence paradigm.[198]

In the product liability context, plaintiffs in defective product cases have used tort law to seek damages from platforms like Amazon.[199] Such actions involved re-tinkering the conception of platforms in a way that imposes liability on them because of their "capacity to situate themselves as a novel form of gatekeeper between third-party suppliers and customers."[200] Despite this, platform governance would rather target the behavior of governed scammers alone—"convenient prox[ies]" that take focus away from the material harms caused by platforms in their expansively designed systems.[201]

Platform governance, or voluntary self-governance, fails to deliver a framework deattenuated from the construct of pseudo-democratically functioning platforms that "do their best" to eliminate platform manipulation. In other words, the platform governance paradigm's core assumption—that social media companies owe a responsibility to "a global network of internet users"—obfuscates the responsibility that platforms owe to their individual users.[202] As a result, platform governance is a hugely unsatisfying paradigm for confronting platform manipulation.

III. PLATFORM DESIGN NEGLIGENCE: A NEW PARADIGM FOR PLATFORM LIABILITY

No present legal paradigm accounts for the deception-related harms that platforms enable against their users. In the wake of this absence, victims and local, state, federal, and even international law enforcers have

---

lobbying, strategic lawlessness, and the unusually low cost of investment capital due to quantitative easing." Id.

197. According to the Yale Law School Information Society Project, "[t]he terms constituting Platform Governance engage with power dynamics and cultural interpretations to create and perpetuate certain technical, political, and legal approaches." Terms and Concepts That Constitute Platform Governance, supra note 192.

198. See infra section III.B (describing the standard of reasonableness that social media companies should meet when designing platforms). The external expectation of reasonableness contravenes the internal self-disciplining expectations that exist within self-governing social media companies.

199. See Catherine M. Sharkey, Products Liability in the Digital Age: Online Platforms as "Cheapest Cost Avoiders", 73 Hastings L.J. 1327, 1329 (2022) ("Judge John Wiley of the California Court of Appeals provocatively described *Loomis*, in which Amazon was held strictly liable for burn injuries caused by a hoverboard listed on its online platform that burst into flames . . . ." (footnote omitted)).

200. Id. at 1344.

201. Id.

202. Terms and Concepts That Constitute Platform Governance, supra note 192.

drawn on an array of methods to address platform manipulation.[203] There is presently no designated civil or criminal enforcement tool that addresses social media companies' liability when malicious actors manipulate their design, resulting in preventable scams and other harms.

Contemporary platform manipulators have managed to evade established American scam policing systems.[204] District attorney's offices and other law enforcement officials are ill-equipped to thread together the large ecosystem of platform-enabled consumer scams.[205] While federal law enforcement has taken action against several platform manipulation schemes, they have thus far been unable to dismantle the multibillion-dollar industry.[206] Time will tell what success, if any, legislative interventions on the table could have on this issue if implemented.[207]

---

203. See Inside the FBI Podcast: Fighting Fraud, FBI, at 3:22 (Aug. 16, 2024), https://www.fbi.gov/news/podcasts/inside-the-fbi-podcast-fighting-fraud (on file with the *Columbia Law Review*) (detailing the FBI's public education methods to combat online scams and the Economic Crimes Unit's role investigating scams by going after wire fraud and mail fraud laws and relying on tips from banks and other information sources); supra sections II.A.–.B.

204. See Lesley Fair, FTC Crunches the 2022 Numbers. See Where Scammers Continue to Crunch Consumers, FTC (Feb. 23, 2023), https://www.ftc.gov/business-guidance/blog/2023/02/ftc-crunches-2022-numbers-see-where-scammers-continue-crunch-consumers [https://perma.cc/5JSZ-L5FM] (describing a thirty percent increase in fraud between 2021 and 2022).; see also Emma Fletcher, Reports of Romance Scams Hit Record Highs in 2021, FTC (Feb. 10, 2022), https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/02/reports-romance-scams-hit-record-highs-2021 [https://perma.cc/WF4M-JDDG] (explaining how "romance scammers are masters of disguise" and that "[m]ore than a third of people who said they lost money to an online romance scam in 2021 said it began on Facebook or Instagram").

205. See Press Release, DOJ, Justice Department Takes Action Against COVID-19 Fraud (Mar. 26, 2021), https://www.justice.gov/opa/pr/justice-department-takes-action-against-covid-19-fraud [https://perma.cc/62ZL-QYTY] (discussing historic enforcement actions against COVID-19-related scammers).

206. See Phil Helsel, Florida Woman Sentenced to 4 Years in Romance Scam that Stole Holocaust Survivor's Savings, NBC News (July 27, 2023), https://www.nbcnews.com/news/us-news/florida-woman-sentenced-4-years-romance-scam-stole-holocaust-survivors-rcna96784 [https://perma.cc/YDC8-FT7G] (describing a scam that targeted the life savings of an eighty-seven-year-old Holocaust survivor); Faith Karimi & Sabrina Souza, Instagram Influencer Scammed Over $2 Million From Older, Lonely Americans, Federal Prosecutors Say, CNN (May 16, 2023), https://www.cnn.com/2023/05/16/us/mona-montrage-alleged-romance-scammer-cec/index.html [https://perma.cc/N9A7-JVYN] (quoting a FBI director as stating that "[r]omance scams . . . are of major concern" (internal quotation marks omitted) (quoting Michael J. Driscoll, Assistant Dir., N.Y. Off., FBI)).

207. See, e.g., Fraud and Scam Reduction Act of 2022, H.R. 1215, 117th Cong. (2022). This bill would have increased governmental efforts to combat and prevent scams that affect seniors, including through the creation of an Office for the Prevention of Fraud Targeting Seniors within the Bureau of Consumer Protection. Id. § 202. Another challenge for legislators is drafting legislation itself; existing fraud statutes are often written too broadly, overly centering the presence of "online hacktivist group[s]" that publish illicitly obtained personal information to the internet. See Philip F. DiSanto, Note, Blurred Lines of Identity Crimes: Intersection of the First Amendment and Federal Identity Fraud, 115 Colum. L. Rev. 941, 950–52 (2015). Importantly, unlike these interventions that would require

Appreciation for the often tacit and menial ways that social media companies design (or fail to design) platforms is essential for imagining a legal regime that begins to impose liability for negligent choices in this burgeoning industry of digital platforms. This Part responds to this challenge by introducing a new paradigm of platform liability, Platform Design Negligence (III.A). This paradigm should inform efforts to combat the novel legal issue of platform manipulation (III.B) and would complement existing legislative and industry reform efforts (III.C).

A. *Platform Design Negligence in Theory*

1. *Overview.* — Legal paradigms reflect images of society that are interpreted by activists, citizens, courts, scholars, and lawyers.[208] The Platform Design Negligence (PDN) paradigm offers a view of law as a system that recognizes the relationship between the holistic design of social media platforms, their architects, and the harms caused by on-platform activity.[209] This paradigm invokes the common law norm of negligence that necessitates four fundamental elements; under this paradigm, victims of platform manipulation can bring a negligence claim if they can establish the following:

> (1) The platform-based company owed them (the platform user) a duty of care;
>
> (2) The company breached that duty;
>
> (3) The breach of that duty caused them some harm; and,
>
> (4) They suffered injuries or damages as a result of that breach.[210]

Applied to platform design, this paradigm tells us that social media companies maintain some degree of liability when they design their platforms in ways that breach their duty to combat platform manipulation. For example, romance scam victims who are extorted by scammers could recover some damages from online dating platforms that recommend scammers as "suggested friends" despite the fact that the online dating platforms knew that the scammers actively maintained multiple profiles,

---

affirmative steps from lawmakers, the Platform Design Negligence paradigm invites courts to apply preexisting negligence principles without the need for legislation. See infra section III.A.

208. See Jürgen Habermas, Paradigms of Law, *in* Habermas on Law and Democracy: Critical Exchanges 13, 13 (Michel Rosenfeld & Andrew Arato eds., 1998) (referring to paradigms as "the background for an interpretation of the system of basic rights").

209. See supra section I.A.

210. According to the foregrounding tort law treatise:

> A person acts negligently if the person does not exercise reasonable care under all the circumstances. Primary factors to consider in ascertaining whether the person's conduct lacks reasonable care are the foreseeable likelihood that the person's conduct will result in harm, the foreseeable severity of any harm that may ensue, and the burden of precautions to eliminate or reduce the risk of harm.

Restatement (Third) of Torts § 3 (Am. L. Inst. 2010).

and had been repeatedly reported for scamming, and yet took no action in response.[211] In this way, the negligence framework resurrects[212] a theory that generates timely consideration of the reputational effects of platform manipulation and produces what law scholars have called "a positive externality in the form of quality information."[213]

Resolving platform manipulation requires moving away from a paradigm of platforms as governors and toward a paradigm of platforms as demystified private actors. While these platforms may have immunity from speech-based torts, they are still liable for how negligently or recklessly designed features create foreseeable and reasonably avoidable injuries.[214] Embracing this new paradigm of PDN requires abandoning the notion of social media platforms as sovereigns, governors, or private "Supreme Courts" with "Oversight Boards,"[215] and instead recognizing platforms as akin to any other company that peddles a product with a design that contributes to harm. Above all, it reflects the current social media landscape, in which new technologies are able to create unprecedented levels of consumer risk "without a corresponding increase in corporate liability."[216] PDN provides recognition for the public rights implicated in social media platforms, which have functionally become digital town squares. To analogize to public nuisance law, PDN embodies the stabilizing effects of tort-based legal liability theories that acknowledge "duties not to interfere with public rights."[217]

---

211. See, e.g., Jim Walsh, Love Hurts: Romance Scam Steals Millions, Sends Burlington County Pair to Prison, Courier Post (Sept. 20, 2024), https://www.courierpostonline.com/story/news/local/south-jersey/2024/09/20/martins-inalegwu-and-steincy-mathieu-get-prison-for-romance-scam/75281131007/ [https://perma.cc/ZD5K-J2G6] (describing how two "[s]cammers struck up relationships on dating websites" and ultimately stole $4.5 million).

212. See Saul Levmore, Richard Posner, the Decline of the Common Law, and the Negligence Principle, 86 U. Chi. L. Rev. 1137, 1155 (2019) (describing the courage of Judge Richard Posner's approach to negligence, which came "a bit too early").

213. Assaf Jacob & Roy Shapira, An Information-Production Theory of Liability Rules, 89 U. Chi. L. Rev. 1113, 1115–18 (2022).

214. See supra section I.B.

215. See Kate Klonick, The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression, 129 Yale L.J. 2418, 2425 (2020) ("Zuckerberg stated in an interview that one could 'imagine some sort of structure, almost like a Supreme Court . . . who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world.'" (quoting Ezra Klein, Mark Zuckerberg on Facebook's Hardest Year, and What Comes Next, Vox (Apr. 2, 2018), https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge [https://perma.cc/W4DR-BDGH])).

216. Rebecca Crootof, The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference, 69 Duke L.J. 583, 589 (2019) (describing how "[internet of things] companies are creating, monitoring, and enforcing contractual-governance regimes with few legal incentives to ensure foreseeable harms are avoided").

217. See Leslie Kendrick, The Perils and Promise of Public Nuisance, 132 Yale L.J. 702, 787 (2023) (arguing for a conception of public nuisance law that recognizes "that we have

PDN is strongly supported by and rooted in the commercial negligence liability paradigm that has evolved within U.S. common law over the past several centuries. Tort law views commercial negligence generally as a function of the corporation's foresight on the harm at issue,[218] with some jurisdictions offering greater deference to public policy considerations.[219] Tort law allows recovery from corporations when they act in this injury-facilitator role by failing to maintain a safe commercial environment or otherwise creating harm-conducive conditions.[220] Thus, PDN calls for an application of this responsibility to the platform economy in a conceptual container for industry, law scholars, and rightsholders alike. It also seeks to provide a structure for defining the duty to exercise reasonable care, which is best assumed by lawmakers.[221]

To illustrate this paradigm, take the hypothetical example of a McDonald's restaurant that opens a brick-and-mortar store that sells coffee. A customer accidentally spills coffee, and the beverage causes third-degree burns on over a fifth of their body, leading to a week-long hospitalization and two years of medical treatment involving skin grafts.[222] Now imagine that a law exists granting restaurants full discretion over the *types* of beverages they sell, but not how they make, sell, and deliver the beverages. Courts proceed to interpret this law to give restaurants like McDonald's full immunity over any harms caused by the temperature of their beverages, what kinds of materials they use for dispensing beverages,

---

duties not to interfere with public rights," what the author calls "a familiar [idea] that has been stigmatized, and at times defanged, in the context of public nuisance through doctrines such as control requirements").

218. Corporate directors and officers are liable to nonshareholder third parties based on their "inadequate management or failure to supervise corporate affairs and subordinates." Martin Petrin, The Curious Case of Directors' and Officers' Liability for Supervision and Management: Exploring the Intersection of Corporate and Tort Law, 59 Am. U. L. Rev. 1661, 1662 (2010); see also In re Caremark Int'l Inc. Derivative Litig., 698 A.2d 959, 970 (Del. Ch. 1996) (holding that boards, regardless of notice, have a duty to ensure reasonable reporting systems).

219. See, e.g., Strauss v. Belle Realty Co., 482 N.E.2d 34, 36 (N.Y. 1985) (finding that it is courts' responsibility "'to limit the legal consequences of wrongs to a controllable degree' . . . and to protect against crushing exposure to liability" (quoting Tobin v. Grossman, 249 N.E.2d 419, 424 (N.Y. 1969))).

220. See Alex Stein, The Domain of Torts, 117 Colum. L. Rev. 535, 549 (2017) (describing how tort law promotes fairness and corrective justice by "allocat[ing] the risks and the costs of accidents").

221. See Mark A. Geistfeld, The Principle of Misalignment: Duty, Damages, and the Nature of Tort Liability, 121 Yale L.J. 142, 149 (2011) (describing the importance of aligning duty in the negligence context to the *class* of cases, categories of actors, patterns of conduct, and other segmenting that allows precise responsiveness to the harms at issue).

222. See, e.g., Retro Report, The Misunderstood McDonald's Hot Coffee Lawsuit, YouTube (Oct. 28, 2019), https://youtu.be/ENTaHxjN4xI?si=M_s0voT1puz_iF0 [https://perma.cc/N8Y3-BQ6Q] (explaining the often misunderstood story of one seventy-nine-year-old woman, one of hundreds burned in that period, who suffered third-degree burns and accrued over $10,000 in medical costs after spilling an extremely hot cup of McDonald's coffee on herself).

the container for dispensing the beverages, and what ingredients they use in their beverages. This interpretation would mark an expansive and illogical extension of the law, obfuscating the nuances of the incremental decisions that restaurants make to create positive experiences for their customers.

In the hypothetical above, now imagine that the customer's injury was directly caused by the actions of a different customer. This malicious customer purposefully stands at the McDonald's "Pick Up" station and shoves customers as they pick up their beverage, causing constant coffee spills and burns for innocent customers. If this incident occurred inside the McDonald's store, and the company agents knew of this issue of actors harming customers as they picked up drinks, and even made it easier for those actors to mistreat customers, McDonald's would be held liable for knowingly and recklessly failing to maintain a safe environment for its customers.

These factors are analogous to the real-world case of McDonald's coffee, in which hundreds of customers were burned by hot coffee and the company refused to act.[223] Eventually a plaintiff sued the company for its negligence and earned a large settlement.[224] The restaurant, like social media companies with internal reporting systems for customers to report suspected platform manipulation, kept an internal log of the incidents but nonetheless failed to act.[225] Social media companies should be similarly liable for platform manipulation harms facilitated by their platform designs.[226]

Applied to social media companies, PDN suggests that social media companies are exposed to tort liability when they (1) design their platforms in ways that they either know or should have reasonably foreseen would create injury and (2) fail to take reasonable action to mediate against the risk created by their platform design. PDN operates the same way as ordinary tort negligence in the context of product liability. When a company creates a heightened risk of harm and fails to act in a reasonable way to address the problem, they are exposed to some degree of liability.[227]

---

223. Id.

224. See Liebeck v. McDonald's Restaurants, P.T.S., Inc., No. CV-93-02419, 1995 WL 360309, at *1 (Dist. Ct. N.M. Aug. 18, 1994) (ordering an award against McDonald's to the Plaintiff "in the amount of $160,000.00 for compensatory damages, and $2,700,000.00 to Plaintiff for punitive damages"), vacated No. CV-93-02419, 1994 WL 16777704, at *1 (Dist. Ct. N.M. Nov. 28, 1994).

225. See Allison Torres Burtka, *Liebeck v. McDonald's*: The Hot Coffee Case, Am. Mus. of Tort L., https://www.tortmuseum.org/liebeck-v-mcdonalds/ [https://perma.cc/QT7L-KGW2] (last visited Feb. 13, 2025) ("The jury learned that 700 other people . . . had been burned before, yet the company did not change its policy of keeping coffee at between 180 and 190 degrees. The company . . . decided that, with billions of cups served annually, this number of burns was not significant.").

226. See Lemmon v. Snap, Inc., 995 F.3d 1085, 1092 (9th Cir. 2021) (finding Snapchat liable for its filter design).

227. See Restatement (Third) of Torts § 3 (Am. L. Inst. 2010).

Section 230, the pinnacle of social media law, does not afford platforms carte blanche over their design choices.[228] The PDN paradigm offers a compatible image of society that strengthens the basis for claims against platform designers by victims of platform designs.

PDN is well supported by state and federal tort law theories of liability. Professor Howard Klemme has offered a "theory of enterprise liability" that is "based on the conviction that underlying the evolutionary development of the common law is an intuitive logic which . . . does exist and is worthy of articulation if possible."[229] PDN carries forth this call by underscoring the conduct social media companies engage in when they *design* their platforms. Other scholars, exploring liability in the design of buildings, have similarly disrupted entity-based theories of liability against building developers by arguing for a liability theory that centers obligations vis-à-vis individual residents.[230] As Judge Guido Calabresi has described, "[T]here is no need for a rigid relation between losses and the scope of the enterprise."[231] Platforms should satisfy their obligations to users to the extent they admit and onboard users to their platforms.

Under PDN, society may begin to appreciate the tremendous magnitude of harms caused by platform manipulation. Victims of platform manipulation may start to understand the multiple vectors through which the social media platforms they use are able to define their experiences. Platforms are well aware of the risk that their products and features can contribute to deception and financial, reputational, and psychological harms. They have the platform design tools to mitigate these harms. Their failure to design their platforms to reasonably address platform manipulation must be scrutinized accordingly. In that analysis, actors—from courts applying common law doctrine to legislators—can begin to fill the gaps of a robust social media platform liability regime.

2. *Platform Design and the First Amendment.* — Similar to § 230, the First Amendment[232] constrains the government's ability to legislate what platforms do, but it does not inoculate platform design from the realm of

---

228. See supra section II.A.

229. Howard C. Klemme, The Enterprise Liability Theory of Torts, 47 U. Colo. L. Rev. 153, 156–57 (1976).

230. See Eric T. Freyfogle, A Comprehensive Theory of Condominium Tort Liability, 39 U. Fla. L. Rev. 877, 879–80 (1987).

231. Guido Calabresi, Some Thoughts on Risk Distribution and the Law of Torts, 70 Yale L. J. 499, 514 (1961).

232. Notably, the First Amendment "permits tailored regulations on employer and employee speech to protect the efficacy of the employment environment and the contrasting rights and dignity of those in it." Francesca Procaccini, Social Network as Work: A Labor Paradigm for Regulating Speech on Social Media, 110 Cornell L. Rev. (forthcoming 2025) (manuscript at 46), https://ssrn.com/abstract=4717216 [https://perma.cc/6LWW-D5AM]. For this reason, a labor paradigm for regulating social media companies may offer a more appropriate application of the First Amendment to social media technology regulation; users provide "labor" to platforms insofar as they input their data. Id.

liability.[233] The First Amendment prohibits Congressional efforts to "abridg[e] the freedom of speech,"[234] meaning that legislative efforts to ascribe liability to social media companies for enabling platform manipulation would only violate the First Amendment if they infringed upon free speech.[235] In 2024, the Supreme Court drew on case law protecting expressive rights of publishers,[236] private utilities,[237] and cable operators[238] to affirm social media companies' ability to exercise discretion over their "prioritization of content," imposition of content labels, and other content moderation practices.[239] Crucially, the Court's extension of First Amendment protection for "how [platform] display[s] [are] ordered and organized"[240] stops at social media "feeds"[241] like Facebook's News Feed tab and YouTube's homepage.[242] While some platform design choices—such as the design of a "feed"—fall under this ill-fated protection, the platform design choices most implicated in platform manipulation do not appear in feeds. Malicious actors can target users by making their own accounts and falsely curating images of legitimacy, accessing the profile pages of other users, direct messaging with targets, and assembling other non-feed displays. Platform design does not necessarily concern itself with users' speech or even the platform's own

---

233. See Genevieve Lakier, The Non-First Amendment Law of Freedom of Speech, 134 Harv. L. Rev. 2299, 2381 (2021) (describing the contamination of free speech discourse by capacious and departmentalist frameworks).

234. U.S. Const. amend. I. ("Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.").

235. The Supreme Court has explicitly held that the First Amendment protects "commercial speech" from companies. Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 566–72 (1980). In *Packingham v. North Carolina*, the Supreme Court stated that "the most important place[] . . . for the exchange of views . . . is cyberspace—the 'vast democratic forums of the internet' in general, and social media in particular." 137 S. Ct. 1730, 1735 (2017) (citation omitted) (quoting Reno v. ACLU, 521 U.S. 844, 868 (1997)).

236. See Moody v. NetChoice, LLC., 144 S. Ct. 2383, 2400 (2024) (citing Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 258 (1974)).

237. See id. (citing Pac. Gas & Elec. Co. v. Pub. Utils. Comm'n, 475 U.S. 1, 12 (1986)).

238. See id. at 2400–01 (citing Turner Broad. Sys., Inc. v. Fed. Commc'ns Comm'n, 512 U.S. 622, 636 (1994)).

239. See id. at 2391 ("Beyond ranking content, platforms may add labels, to give users additional context. And they also remove posts entirely that contain prohibited subjects or messages, such as pornography, hate speech, and misinformation on certain topics. The platforms thus unabashedly control the content that will appear to users.").

240. Id. at 2406.

241. See Klonick, New Governors, supra note 128, at 1660 (describing how content is displayed specifically on "newsfeed[s], homepage[s], or stream[s]").

242. See *NetChoice*, 144 S. Ct. at 2406 ("The current record suggests the opposite as to Facebook's News Feed and YouTube's homepage. When the platforms use their Standards and Guidelines to decide which third-party content those feeds will display, . . . they are making expressive choices. And because that is true, they receive First Amendment protection.").

speech.[243] Definitionally, platform design choices function as ruled lines on a sheet of paper—they certainly inform the users' speech experience but do not necessarily cross the threshold of abridging the freedom of speech.

B. *The Platform Design Negligence Paradigm in Practice*

The PDN paradigm stands for the proposition that social media companies are directly responsible to each individual user, and when those companies make design choices that facilitate deception through their platforms, they may be negligent. A number of courts have recognized that platform design decisions do not receive § 230 immunity.[244] At the same time, courts, law enforcers, and plaintiffs alike struggle to point to common law or statutory bases for their arguments linking their harms to the platform design choices.[245] PDN represents an entry point for lawmakers and industry professionals seeking to curb platform manipulation on their platforms.[246] It operates on a dual track, first drawing on background presumptions of tort liability under federal and state common law to bring PDN claims, and, second, guiding lawmakers to pass legislation that prescribes social media liability for platform manipulation and shields PDN claims from arbitration agreements, among other measures.[247]

Tort law is a powerful tool for holding corporate actors accountable when they themselves do not engage in the primary conduct that causes injury to customers, but they nonetheless contribute to the injury.[248]

---

243. Even if so, commercial speech doctrine would fail to shield social media companies from regulation targeting platforms' misleading or deceptive designs because the "speech-design" that exposes users to heightened risk of scam and other deception falls squarely within Congress' jurisdiction. See Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 566 (1980) (outlining a four-step test that asks whether commercial speech "concern[s] lawful activity and [is] misleading").

244. See Forrest v. Meta Platforms, Inc., 737 F. Supp. 3d 808, 818 (N.D. Cal. 2024) (denying Meta's § 230 affirmative defense when Meta contributed to the appearance of scam ads).

245. See, e.g., Roland v. Letgo, Inc., 644 F. Supp. 3d 907, 917 (D. Colo. 2022) ("Plaintiffs have not cited a single case in which a court held an internet platform potentially liable for violent criminal acts perpetrated by a platform user who lured an innocent consumer into a scheme through means of misrepresentations made by the criminal.").

246. For a discussion on entry points for lawmakers, see infra section III.C.3.

247. See supra note 91. Arbitration agreements increasingly play dangerous roles in consumer law, and PDN requires the exact litigation pathway that arbitration agreements have been interpreted to foreclose. See David Horton, Arbitration About Arbitration, 70 Stan. L. Rev. 363, 377–99 (2018) (discussing the modest ambitions of the Federal Arbitration Act, which "abrogated hundreds of years of common law").

248. While tort law offers a helpful framework for discussing platform design, it is not a be-all and end-all solution. Tort law is inadequate at addressing nonfinancial injuries such as the economic and information-based injuries at the heart of platform manipulation. See Rustad & Koenig, supra note 83, at 1482 ("Since the enactment of section 230, no U.S. court has recognized or upheld a judgment against a social media provider arising out of third-

Though tort law has failed to rein in harmful corporate conduct in industries it clearly applies to, including the automotive, aerospace, and consumer chemical industries, PDN circumvents those shortcomings associated with undocumented relational lines between the harm at issue and effects.[249] Because platform manipulation exists directly on platforms and platform designs are visible to the lay user, platform manipulation's contours are more readily visible under PDN; platforms are privy to the ways their designs are exploited.[250] Though central regulation that prevents manipulation before it occurs would maximize consumer protectionism, PDN's construction of platform design presents a baseline for realizing a comprehensive regulatory regime to effectively regulate social media. Moreover, PDN is practical because judges can apply it under existing principles, meaning that it is available immediately, and federal legislation has thus far failed to materialize on this issue.

Individual social media users could prove harm under PDN in a variety of ways. Scam victims can argue that platforms failed to take reasonable measures against designing the platforms in ways that augment, accelerate, and accredit scammers. For investment, job, romance, and similar scams involving fund transfers, the financial harm will involve a complex inquiry that apportions loss pursuant to the time-tested joint and several liability common law doctrines.[251] Under PDN, plaintiffs could also pursue remedies for emotional harm, psychological harm, lost time, lost political power, and communal harms, through personal testimony, expert testimony, scam experts, psychologists, witness statements, research and data on scam impacts, and more.

Liability for platform design also provides deterrent effects for the industry, incentivizing improved platform design and the development of rigorous investments in anti-scam features, as have been adopted in the disinformation, misinformation, and AI-generated deepfake contexts.[252]

---

party publication torts on a social network."); see also Leonard J. Feldman & Julia Doherty, The Class of Injuries Test: A Unifying Proposal to Determining Duty, Proximate Cause, and Superseding Cause in Negligence Claims, 47 Seattle U. L. Rev. 1613, 1621 (2024) (discussing difficulties with applying foreseeability principles in third-party contexts).

249. See generally Bryan H. Choi, Crashworthy Code, 94 Wash. L. Rev. 39 (2019) (describing how the "crashworthy" liability doctrine, which holds corporations liable for their unsafe designs that lead to harm, was developed in response to automobile accidents but is better applied in the software context).

250. Under a consumer protectionist lens, "burdens caused by new technologies should not be forced upon hapless victims, but should be borne instead by those best situated to account for those risks." Id. at 50.

251. See Nancy C. Marcus, Phantom Parties and Other Practical Problems With the Attempted Abolition of Joint and Several Liability, 60 Ark. L. Rev. 437, 438–44, 484–86 (2007) (describing the challenges with fault allocation systems and arguing that pure joint and several liability paired with contribution can best serve the aims of tort law).

252. See Hayden Field, Tech Layoffs Ravage the Teams that Fight Online Misinformation and Hate Speech, CNBC (May 26, 2023), https://www.cnbc.com/2023/05/26/tech-companies-are-laying-off-their-ethics-and-safety-teams-.html [https://perma.cc

Platforms have already developed extensive tools for monitoring, detecting, and combating disinformation and misinformation: They track malicious actors, label them, and remove them from the platform.

PDN also speaks to the ambiguity left in the wake of *Twitter, Inc. v. Taamneh*, in which the Supreme Court determined that plaintiffs failed to show that a social media platform's algorithmic choices rose to the level of impermissible conduct under the Justice Against Sponsors of Terrorism Act (JASTA).[253] There, the conduct at issue was highly attenuated insofar as the plaintiffs could not connect the real-world terrorist attack with the terrorist group's use of social media.[254] On the other hand, in platform manipulation, individual social media users are victimized by the on-platform conduct that serves as the basis of the PDN claim.[255] Platform design operates as customer service—the principal business relationship that the Supreme Court in *TransUnion LLC v. Ramirez* found that concrete injuries in fact arose from.[256] Specifically, the Supreme Court affirmed the presence of concrete injuries when customers' platforms were tainted by misleading statements (i.e., labels) imposed on the customer's profile and exported to third parties.[257] Platform manipulation more clearly creates real-world harms to victims than did the credit check company's wrongful labeling of customers as "terrorists" in *TransUnion*; the harm to customers in platform manipulation bears "a 'close relationship' to the harm" that is already recognized in tort liability for consumer product designs.[258] Thus PDN claims are ripe for success under the current standard for proving standing with monetary and nonmonetary injuries—claims that when properly brought under the "typical limits on tort liability" could affect industry incentives.[259]

Platforms can enhance disclaimers or notifications in messaging features to advise users when they are messaging with other users who have been repeatedly reported for consumer scams. They can monitor users who are sending hundreds of messages to strangers a day. They can use metadata to flag and isolate spam actors. Platforms can also engage in anti-addiction platform design that limits the harms caused by addictive design

---

/2MYM-JD79] (discussing how several platforms conducted layoffs in 2023 on teams that worked on platform manipulation).

253. 143 S. Ct. 1206, 1230–31 (2023).

254. Id. at 1227–28 ("Plaintiffs do not claim that defendants intentionally associated themselves with ISIS' operations or affirmatively gave aid that would assist each of ISIS' terrorist acts. Nor have they alleged that defendants and ISIS formed a near-common enterprise of the kind that could establish such broad liability.").

255. Cf. id. at 1228 ("These allegations are thus a far cry from the type of pervasive, systemic, and culpable assistance to a series of terrorist activities that could be described as aiding and abetting each terrorist act.").

256. 141 S. Ct. 2190, 2208–09 (2021).

257. Id.

258. Id. at 2209.

259. *Taamneh*, 143 S. Ct. at 1228–29.

features;[260] they can also use notification systems as a dimension for policy interventions. These processes could be replicated to combat platform-enabled consumer scams. Platforms could also develop proactive detection mechanisms that actively discover helpful signals for identifying accounts that pursue platform-enabled scams. This detection could transfer to labels.

With the PDN paradigm in practice, platforms would better understand when they face liability: when they understand the risk, fail to act, and reasonably could design their platforms alternatively. Federal and state lawmakers can provide legislation that describes "reasonable platforms." The paradigm could also incentivize or require platforms to invest in content moderation systems that provide protections for those most vulnerable to online scams[261] and build capacity in a wider range of demographics, dialects, and regions. For example, the lack of investment in content moderation systems that address a wide range of demographics has been linked to the proliferation of violent and extremist content.[262] Similar investments in monitoring capacities for scam content could assist efforts to identify worldwide networks of scammers on social media platforms.

Unlike the platform governance paradigm that treats platforms like government entities, the PDN paradigm situates platforms like private corporations. When they decide to design their platforms to invite abuse and deception, they operate like an amusement park that uses poor architecture to design unsafe rides. This concept can help clarify the bounds of reasonable and unreasonable behavior on the part of lawmakers, social media companies, and legal thinkers alike.

Red team exercises, a type of alternative analysis or stress testing[263] that is increasingly prevalent in the AI governance field, could be

---

260. See Press Release, Eur. Parliament, New EU Rules Needed to Make Digital Platforms Less Addictive (Oct. 25, 2023), https://www.europarl.europa.eu/news/en/press-room/20231023IPR08161/new-eu-rules-needed-to-make-digital-platforms-less-addictive [https://perma.cc/L32Y-QFLC] (describing the European Parliament's demand for nonaddictive platform designs such as "turning off notifications by default; chronological feeds; greyscale mode; warnings or automatic locks after a pre-set time use," and more).

261. See Ctr. for Countering Digit. Hate, Deadly By Design 24 (2022), https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf [https://perma.cc/8QLA-EMS4] (describing how some users are more vulnerable to certain kinds of platform manipulation than others).

262. See, e.g., Faiza Patel & Laura Hecht-Felella, Facebook's Content Moderation Rules Are a Mess, Brennan Ctr. for Just. (Feb. 22, 2021), https://www.brennancenter.org/our-work/analysis-opinion/facebooks-content-moderation-rules-are-mess [https://perma.cc/6KJD-4VND] (linking Facebook's content moderation decisions to conflict in the Nagorno-Karabakh region, among other cases in which Facebook tools "fail to adequately account for context or pollical, cultural, linguistic, and social differences").

263. See Rory Van Loo, Stress Testing Governance, 75 Vand. L. Rev. 553, 557 (2022) (arguing that "well-designed stress tests can provide Congress with a mechanism to supervise agencies' readiness to safeguard society").

exceptionally fruitful both for platforms and legal scholars[264] looking for guidance.[265] In one case, red team exercises exposed X's neglect of the potential for child sexual exploitation that would result from a design choice—creating a new account type that would be permitted to monetize adult content.[266]

For courts, victims, and law enforcers, this paradigm presents the opportunity to revisit and shine new light on previous cases in which legal frameworks failed to account for the exceptional role of platform design in platform manipulation. For example, in the case of *Doe v. Grindr Inc.*, when a district court rejected negligence and product liability claims brought by a fifteen-year-old who was sexually assaulted by sexual predators he met on the online dating platform Grindr, an eye towards platform design could have yielded a different result for the victim.[267] There, the District Court for the Central District of California determined that the platform's decisions to create "matches" and offer minimal age-verification procedures failed to implicate § 230.[268] Under PDN, the plaintiff may have considered an alternative series of claims to vindicate his rights against the platform for its role in his victimization. Claims of actions against Grindr for its negligence in failing to design controls that could have limited the age groups with which the fifteen-year-old could have been matched with would have likely survived scrutiny under the paradigm and existing laws.

C. *Legislative Reforms and Industry Solutions*

One way to actuate the PDN paradigm is for states to effectuate their own existing or forthcoming tort laws to clarify the bounds of reasonableness in platform design. Judges can interpret existing tort laws

---

264. See Miles Brundage et al., Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims 2 (Apr. 2020) (unpublished manuscript), https://arxiv.org/pdf/2004.07213 [https://perma.cc/P46L-S9QN] (detailing evidence-backed mechanisms for enhancing safety in AI systems).

265. See Stacy-Ann Elvy, Paying for Privacy and the Personal Data Economy, 117 Colum. L. Rev. 1369, 1376–77, 1459 (2017) (explaining the complexity associated with tackling digital harms in power-imbalanced relationships with platforms).

266. See Zoë Schiffer & Casey Newton, How Twitter's Child Porn Problem Ruined Its Plans for an OnlyFans Competitor, The Verge (Aug. 30, 2022), https://www.theverge.com/23327809/twitter-onlyfans-child-sexual-content-problem-elon-musk [https://perma.cc/9N63-8BTK] (describing the effects of red team exercises on mitigating online harms).

267. See Doe v. Grindr Inc., 709 F. Supp. 3d 1047, 1050–51, 1054–55 (C.D. Cal. 2023) (finding that the defective product design claims, among others, were barred by § 230 immunity).

268. Id. at 1057 (stating that "Section 230 immunizes Grindr from Doe's claims" particularly because "[Doe's] allegations suggest only that [Grindr] 'turned a blind eye' to the unlawful content posted on its platform, not that it actively participated in sex trafficking" (alterations in original) (internal quotation marks omitted) (quoting Does 1–6 v. Reddit, Inc., 51 F.4th 1137, 1145 (9th Cir. 2022))).

to apply to platform design without touching § 230.[269] Congress should also enact a tort law statute on the matter. Congress has previously enacted tort law statutes, like the Alien Tort Statute that applies to private defendants[270] and the Federal Tort Claims Act that allows plaintiffs compensation from the U.S. government.[271] Crucially, a federal platform design statute must exempt these claims from arbitration agreements in cases when the network effects and power imbalance create distressing social harm in the form of successful scams.[272]

Tort law is a logical choice for victims of scams on social media. Its focus on harm lends itself to applications in the context of harm inflicted through the internet. Corporate liability jurisprudence seems to be headed in this direction; notably, personal injury attorneys specializing in tort law have been able to achieve historic wins in the gun product liability context.[273] Platforms already have a duty to warn when they hold information obtained from an outside source about a scheme on their platforms.[274] Tort case law on platform manipulation issues is highly sparse and ripe for innovation. For example, banks are unlikely allies insofar as they can bring PDN claims against the social media companies that act as "first responders" for many scams and other fraudulent activity.[275]

State lawmakers have an instrumental role to play as well. Some proposals call for slowing down transfers to mitigate the financial harm of scams.[276] One legislator has introduced a multifaceted plan to confront

---

269. See supra section II.A.

270. 28 U.S.C. § 1350 ("The district courts shall have original jurisdiction of any civil action by an alien for a tort only, committed in violation of the law of nations or a treaty of the United States.").

271. See Michael D. Contino & Andreas Kuersten, Cong. Rsch. Serv., R45732, The Federal Tort Claims Act (FTCA): A Legal Overview 1 (2023), https://crsreports.congress.gov/product/pdf/R/R45732 [https://perma.cc/FD64-LUWA] (explaining broadly written statutes that permit tort claims by non-U.S. citizens and torts committed by U.S. employees).

272. See Horton, supra note 247, at 440 (highlighting that "companies are attempting to privatize [the courts'] gatekeeping function"); supra note 91 and accompanying text (highlighting the vulnerabilities of social media ToS agreements).

273. See Michael Steinberger, The Lawyer Trying to Hold Gunmakers Responsible for Mass Shootings, N.Y. Times (Sept. 29, 2023), https://www.nytimes.com/2023/09/29/magazine/the-lawyer-trying-to-hold-gunmakers-responsible-for-mass-shootings.html (on file with the *Columbia Law Review*) (describing the wrongful death tort lawsuit that pierced perceived statutory immunity for gun manufactures to hold accountable a gun company that dangerously marketed its goods).

274. See Doe v. Internet Brands, Inc., 824 F.3d 846, 850–51 (9th Cir. 2016) ("California law imposes a duty to warn a potential victim of third-party harm when a person has a 'special relationship to either the person whose conduct needs to be controlled or . . . to the foreseeable victim of that conduct.'" (quoting Tarasoff v. Regents of Univ. of Cal., 551 P.2d 334, 343 (Cal. 1976), superseded by statute, Cal. Civ. Code § 43.92 (2013))).

275. See supra note 20.

276. Ryan Sabalow, A California Senior Lost $700k to Scammers. Newsom Rejected Bill to Slow Bank Transfers, Cal Matters (June 19, 2024), https://calmatters.org/digital-democracy/2024/06/california-senior-fraud-scam/ [https://perma.cc/RS48-CJ93] (last updated Sept. 28, 2024).

platform manipulation that targets elders in New Jersey, including through the Empowering States to Protect Seniors Against Bad Actors Act, which would potentially build anti-scam enforcement capacity.[277]

In the meantime, Congress is occupied with a narrower set of issues. A handful of lawmakers "are now looking to defamation law as a social fix for systemic problems rather than a remedy for harm to individual reputation."[278] In 2023, the Preventing Deepfakes of Intimate Images Act was introduced to criminalize the sharing of nonconsensual images and sexually explicit AI-generated content.[279] Later, in 2024, the Disrupt Explicit Forged Images and Non-Consensual Edits Act was introduced to provide a cause of action for the creation and distribution of "digital forgery" when the victim had not given consent.[280] Proposed legislative interventions face an unknown fate. While deepfakes, sextortion, and similar crimes garner attention, they do not account for vast majority of platform-enabled scams at play in the United States.[281]

U.S. federal lawmakers have offered a few other legislative solutions to the issue of platform manipulation, though none of these address platform design. The Fraud and Scam Reduction Act would hone in on scams targeting elders by establishing a new advisory group and office within the FTC.[282] This Act would create a system of voluntary agreements and partnerships with social media companies[283] even though behavioral remedies such as platform design enhancements could play a superior role. Importantly, these voluntary public–private coalitions fail to create anything proximate to a private right of action or civil enforcement vessel for victims of platform manipulation.

CONCLUSION

As platform manipulators develop increasingly sophisticated methods for exploiting social media to serve their malicious objectives, victims of

277. Press Release, Josh Gottheimer, As Part of Senior Security Strategy, Gottheimer Announces New Action to Combat Senior Scams on Social Media and More (May 6, 2024), https://gottheimer.house.gov/posts/release-as-part-of-senior-security-strategy-gottheimer-announces-new-action-to-combat-senior-scams-on-social-media-and-more [https://perma.cc/JT5F-CP4C].

278. Lili Levi, Disinformation and the Defamation Renaissance: A Misleading Promise of "Truth", 57 U. Rich. L. Rev. 1235, 1240 (2023).

279. Preventing Deepfakes of Intimate Images Act of 2023, H.R. 3106, 118th Cong. § 1 (2023).

280. Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024, H.R. 7569, 118th Cong. § 1 (2024).

281. See supra Part I.

282. Fraud and Scam Reduction Act of 2022, H.R. 1215, 117th Cong. § 102 (2022). The bill would have increased governmental efforts to combat and prevent scams that affect seniors, including through the creation of an Office for the Prevention of Fraud Targeting Seniors within the Bureau of Consumer Protection. See supra note 207.

283. Id.

this activity, legal actors, and social media companies will continue to pursue measures that prevent and respond to these harms. Victims will continue to seek justice, as plaintiffs and law enforcers pursue action on their behalf. The current frameworks for confronting the harms perpetuated by platform manipulation fail to adequately account for platform design as a vector of chargeable conduct.

Platform Design Negligence is a container for articulating future possibilities at the crossroad between private power and the law. In this universe, the public does not view platform executives as mere governors of social media. Rather, the public recognizes the platform's duties to users. Social media companies that take steps to enable platform manipulation through their tacit and understated toolkit—platform design—must begin to face the music whenever their choices contribute to real-world harm.