"A Change Is Gonna Come:" Developing a Liability Framework for Social Media Algorithmic Amplification

Amy B. Cyphert

Jena Martin
St. Mary's University School of Law, jmartin9@stmarytx.edu

"A Change is Gonna Come:"[1] Developing a Liability Framework for Social Media Algorithmic Amplification

Amy B. Cyphert* & Jena T. Martin**

From the moment social media companies like Facebook were created, they have been largely immune to suit for the actions they take with respect to user content. This is thanks to Section 230 of the Communications Decency Act, 47 U.S.C. § 230, which offers broad immunity to sites for content posted by users. But seemingly the only thing a deeply divided legislature can agree on is that Section 230 must be amended, and soon. Once that immunity is altered, either by Congress or the courts, these companies may be liable for the decisions and actions of their algorithmic recommendation systems, artificial intelligence models that sometimes amplify the worst in our society, as Facebook whistleblower Frances Haugen explained to Congress in her testimony.

But what, exactly, will it look like to sue a company for the actions of an algorithm?

Whether through torts like defamation or under certain statutes, such as those aimed at curbing terrorism, the mechanics of bringing such a claim will surely occupy academics and practitioners in the wake of changes to Section 230. To that end, this Article is among the first to examine how the issue of algorithmic amplification might be addressed by agency principles of direct and vicarious liability, specifically within the context of holding social media companies accountable. As such, this Article covers the basics of algorithmic recommendation systems, discussing them in layman's terms and explaining why Section 230 reform may spur claims that have a profound impact on traditional tort law. The Article looks to sex trafficking claims made against social media companies—an area already exempted from Section 230's shield—as an early model of how courts might address other claims against these companies. It also examines the potential hurdles, such as causation, that will remain even when Section 230 is amended. It concludes by offering certain policy considerations for both lawmakers and jurists.

[1] The title is taken from the song "A Change is Gonna Come" written by Sam Cooke.

* Lecturer in Law and Director, ASPIRE Office, West Virginia University.

** Professor of Law, West Virginia University. This article was supported, in part, by Hodges Fund Faculty Research Grants. The authors are grateful to the following colleagues and friends for their feedback: Valarie Blake, Caroline Osborne, Sam Perl, John Taylor, and Kirsha Trychta. In addition, Jamie Zepeda, Sophia Runion, and Valeria Lopez-Altamirano provided dedicated research support. The authors are grateful to the editors of the *UC Irvine Law Review* for their excellent editorial assistance.

INTRODUCTION

"[*Facebook*] *has created algorithms that are deadly good at pointing you toward everyone, everywhere who offers more of what you seem to want* [*and*] *has utterly failed to* *use those same data/algorithms to back up their own claimed child safety policies."* *—Prof. Lara Putnam*[2]

As part of her work with the University of Pittsburgh's Disinformation Lab, Professor Lara Putnam has spent a lot of time on Facebook[3] examining private groups.[4] In October 2021, she noticed a disturbing phenomenon: a number of Facebook groups with names like "looking for a boyfriend/girlfriend who is 9,10,11,12 [sic] or 13 years old" were flourishing, garnering thousands of followers.[5] Both the content and the images appearing on the threads of these public forums violated Facebook's Community Standards[6] (including images stating "Xxx Porno chicas").[7] However, Prof. Putnam's attempt to report these groups has been met with only marginal success. In most instances, Facebook responded to Prof. Putnam's reports by stating that they had "reviewed" the content and found it to meet the company's standards (even though those same standards[8] provide that children under 13 should not be using the site).[9] Many of these groups are based in Latin American and African countries—both places that have been labeled as "the rest of the world" in leaked Facebook documents,[10] and as such, these groups are

[2] Lara Putnam (@lara_putnam), TWITTER (Nov. 30, 2021, 6:43 AM), https://twitter.com/lara_putnam/status/1465692986331721734 [https://web.archive.org/web/20211202014423/https://twitter.com/lara_putnam/status/1465692986331721734]. Professor Putnam is a faculty member at the University of Pittsburgh and is the Co-Lead for the Southwest Pennsylvania Civic Resilience Initiative of the Pitt Disinformation Lab, located in the university's Institute for Cyber Law, Policy, and Security.

[3] In 2021, Facebook officially changed the name of its parent company to Meta. We use Facebook throughout this Article to refer to the social media site as well as the corporation now known as Meta.

[4] Putnam, *supra* note 2.

[5] Lara Putnam (@lara_putnam), TWITTER (Nov. 11, 2021, 7:55 PM), https://twitter.com/lara_putnam/status/1459007002340995076 [https://web.archive.org/web/20211112035601/https://twitter.com/lara_putnam/status/1459007002340995076]; Lara Putnam (@lara_putnam), Twitter (Nov. 12, 2021, 5:31 PM), https://twitter.com/lara_putnam/status/1459333042472988674 [https://web.archive.org/web/20220119090714/https://twitter.com/lara_putnam/status/1459333042472988674].

[6] *Facebook Community Standards*, META, https://transparency.fb.com/policies/community-standards/ [https://perma.cc/M72L-FSN7] (last visited Oct. 20, 2022).

[7] Lara Putnam (@lara_putnam), TWITTER (Nov. 21, 2021, 7:12 PM), https://twitter.com/lara_putnam/status/1462620022854107145 [https://web.archive.org/web/20211122054644/https://twitter.com/lara_putnam/status/1462620022854107145].

[8] *How Do I Report a Child Under the Age of 13 on Facebook?*, META, https://www.facebook.com/help/157793540954833 [https://perma.cc/GA7W-QRLY] (last visited Oct. 20, 2022).

[9] Lara Putnam (@lara_putnam), TWITTER (Nov. 11, 2021, 7:57 PM), https://twitter.com/lara_putnam/status/1458886574301712395 [https://web.archive.org/web/20220723194359/https://twitter.com/lara_putnam/status/1458886574301712395].

[10] Cat Zakrzewski, Gerrit De Vynck, Niha Masih & Shibani Mahtani, *How Facebook Neglected the Rest of the World, Fueling Hate Speech and Violence in India*, WASH. POST (Oct. 24, 2021, 7:00 AM), https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech [https://perma.cc/F6JS-JJTK] ("[Facebook's] budget to fight misinformation was heavily weighted toward America, where 84 percent of its 'global remit/language coverage' was allocated. Just 16 percent was earmarked for the 'Rest of World,' a cross-continent grouping that included India, France and Italy.").

subject to significantly less human oversight and review with respect to the moderation of the content in the groups.[11]

Nonetheless, at least *some* of Facebook's algorithms seem to be working as intended.

Although Facebook uses machine learning algorithms to power both its content moderation practices and its recommendation system,[12] as Prof. Putnam's experience documents, there are significant problems with both of these systems, including that they do not appear to speak to each other. Despite Prof. Putnam repeatedly flagging content as objectionable, Facebook's recommendation algorithm suggests that she join a stream of similar groups—groups targeting young children and having the hallmarks of a trafficking scheme.[13]

Sadly, the use of Facebook and other social media sites to further human trafficking schemes is not the only way these sites can cause harm. In fact, large amounts of violent, criminal, and otherwise disturbing content has been disseminated using social media sites. Changes made to Facebook's newsfeed algorithms in approximately 2018[14] caused divisive and enraging content to be seen by even *more* people. This effect is partially the product of a decision by Facebook to prioritize increases in user engagement over other objectives.

And yet, Facebook has been largely immunized for any role it might have played in causing harm. Why? Courts have uniformly held that a provision in the Communications Decency Act—Section 230—provides a powerful and near-absolute liability shield to Facebook and other social media actors for content posted by others to their sites.[15] Notably, Section 230 was enacted at a time before social media was even in existence. Now, it has come to dominate the issue of who can be liable for the spread of misinformation, disinformation, and violent content. While the jurisprudence construing Section 230 has almost always found it to be an absolute bar to social media company liability, change is clearly on the horizon. Several bills on the topic are now pending before Congress, and Section 230 reform is one of the only things both Democrats and Republicans seem able to agree upon.[16] The testimony of Facebook whistleblower Frances Haugen has only accelerated that process. Further, Supreme Court Justice Thomas has all but invited

[11] *Id.*

[12] With Facebook's recommendation system, users are directed to other users they may wish to connect with, or have certain material placed at or near the top of their news feeds, a process also known as "algorithmic amplification."

[13] *See, e.g.*, Christopher Ljundquist, *Sex Trafficking of Minors: Know the Process, Look for the Signs*, BRIDGING REFUGEE YOUTH & CHILD.'S SERVS., https://brycs.org/anti-trafficking/sex-trafficking-of-minors-know-the-process-look-for-the-signs/ [https://perma.cc/332R-HTMW] (last visited Oct. 20, 2022). It is especially puzzling that the groups that Prof. Putnam has flagged appear to remain on Facebook given that claims of sex trafficking are removed from Section 230's liability shield, as discussed in Part III *infra*.

[14] *See* Part I *infra*.

[15] *Id.*

[16] *See* Section II.A *infra*.

attorneys to challenge the judicial precedent that gives such expansive reach to the liability shield,[17] and in October of 2022, the Supreme Court granted cert in just such a case.[18] As such, it seems highly likely that Facebook and other social media companies will, in some way, lose the shield that Section 230 provides.

However, many of the actions that might give rise to liability for companies like Facebook result from a complicated series of algorithmic decisions (rather than human intermediaries). So, the question becomes, what's next for the law in holding these companies liable? In short, what approaches will plaintiffs use to hold companies liable for the harm caused by their algorithms?

This Article aims to answer that question, as well as to look at some of the likely hurdles that plaintiffs will continue to face even in the absence of Section 230.[19]

Using principles of agency law as well as the development of corporate accountability theories under human rights law, we analyze how corporate liability might evolve (or be established)[20] when a company uses algorithms to display injurious content.[21] Specifically, we conclude that, absent a major overhaul to the

[17] Malwarebytes, Inc. v. Enigma Software Group USA, LLC, 141 S. Ct. 13, 18 (2020) (mem.) (Thomas, J., concurring), *denying cert. to* 946 F.3d 1040 (9th Cir. 2019).

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

[19] This Article builds on the work of other scholars who have attempted to create frameworks for addressing the legal challenges of artificial intelligence/machine learning. *See, e.g.*, Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829 (2019) (advocating for a judicial framework to assess explainable AI decisions); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117 (2014) (discussing a products liability framework for autonomous machines like self-driving cars). Indeed, many articles (such as ours) have even wrestled with tort and agency law principles in trying to determine an appropriate framework. *See, e.g.*, Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353 (2016) (discussing the potential costs and benefits of government regulation of AI); Iris H.-Y. Chiu & Ernest W.K. Lim, *Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm*, 20 WASH. U. GLOBAL STUD. L. REV. 347 (2021) (discussing the use of machine learning algorithms with a corporate framework); Pinchas Huberman, Essay, *Tort Law, Corrective Justice and the Problem of Autonomous-Machine-Caused Harm*, 34 CAN. J.L. & JURIS. 105 (2021); Marta Infantino & Weiwei Wang, *Algorithmic Torts: A Prospective Comparative Overview*, 28 TRANSNAT'L L. & CONTEMP. PROBS. 309 (2019) (discussing the fault-based frameworks for liability); Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315 (2020) (discussing the complications that arise in applying a negligence framework to AI).

[20] Other countries have already begun grappling with the implications of liability in an AI world. *See, e.g.*, Stefan Heiss, *Towards Optimal Liability for Artificial Intelligence: Lessons from the European Union's Proposals of 2020*, 12 HASTINGS SCI. & TECH. L.J. 186 (2021) (discussing the EU's proposal for a liability regime for AI). However, plaintiffs in other countries do not face the additional hurdle of Section 230 that we address here.

[21] Although we are using Facebook in particular as a lens for this analysis, we believe that the framework we propose herein will be helpful for advocates who seek to hold companies of all kinds responsible for the actions of their algorithms. Specifically, this Article examines liability for the concept of algorithmic amplification.

legal framework (which one of us argues elsewhere is needed),[22] agency law concepts can be a useful stopgap to help establish these principles. As such, this Article proceeds as follows: In Part I, we provide context for the controversy by examining social media companies' use of algorithms to both moderate and amplify content. Using Facebook as a case study, we discuss how these algorithms—which are intentionally[23] designed to maximize engagement thereby increasing corporate profits—methodically feed their users information that sometimes leads to tragic results. In Part II, we analyze the changes we see coming, likely a combination of changes (made by Congress or by courts) to Section 230 as well as other legislative proposals that would provide more regulation and oversight of social media companies. We specifically discuss the particular set of circumstances that have unfolded to allow either Congress or the courts to roll back the putative prohibition in Section 230 that has prevented harmed users and their families from bringing a successful suit against social media companies. However, because many of the controversial decisions in question are being executed by algorithms rather than humans, any actions brought after the removal of Section 230's liability shield must grapple with how to hold a company liable for its algorithmic system. As such, in Part III we provide our analytical framework on how plaintiffs may craft these claims and how courts might respond. Specifically, in this Section we discuss agency law principles of vicarious and direct liability and examine a number of potential claims that victims could proffer against social media sites. In Part IV, we offer our discussion of policy considerations for lawmakers and courts to guide them in their analysis of the issue.

One way or another, a "change is gonna come" for Section 230. This Article provides some insight into the practical impact that change will have.

I. FACEBOOK, CONTENT MODERATION, AND ALGORITHMIC AMPLIFICATION

Social media companies employ a range of practices and tools to moderate the content on their sites. Sometimes this moderation is what we think of as "traditional" content moderation, such as taking down objectionable content. Other times, this content moderation involves amplifying content that is more likely to engage users, through positioning a post at the top of a user's newsfeed, for example, or recommending that the user follow other users. Although both

[22] Jena Martin & Lara Putnam, The Shrinking Horizons of a Brave New (Digital) World (unpublished manuscript) (on file with author).

[23] While we discuss intentionality here within the context of human designers, there are others who argue for a use of intentionality within the AI framework itself. *See, e.g.*, Mihailis E. Diamantis, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, 98 N.C. L. REV. 893 (2020) (arguing for the extension of legal concepts such a "knowing" and "deliberately" to machine-learning algorithms). For an insightful discussion of whether algorithms can be trusted, see David Spiegelhalter, *Should We Trust Algorithms?*, HARV. DATA SCI. REV. (May 23, 2022, 10:32 AM), https://hdsr.mitpress.mit.edu/pub/56lnenzj/release/1?utm_campaign=Jeroen%20Verkroost%27s%20Newsletter&utm_medium=email&utm_source=Revue%20newsletter [https://perma.cc/KT5Z-NU5W].

practices can be termed "content moderation," they are quite different and the incentives the companies have to perform them are also different. Facebook provides a helpful lens for exploring these practices, since it is the largest social media company in the world (executives note that one-third of the world's population is on the company's platforms),[24] and therefore has billions of pieces of content posted to it each day, providing huge content moderation challenges. Facebook is also a good lens through which to examine content moderation since it publicly provides some information about its content moderation practices[25] (though, as will be explored below, there is evidence suggesting that the company is not always following its own public statements about content moderation, and the company is much less transparent about its algorithmic amplification practices).

A. Traditional Content Moderation at Facebook

Like all major social network companies, Facebook uses a combination of humans and artificial intelligence to perform content moderation.[26] Predictive algorithms fueled by machine learning[27] perform the first level of moderation, automatically removing certain posts and flagging others based on keywords, user flagging, and other data.[28] At Facebook, the decisions surrounding what content is allowed on the platform are guided by the platform's Community Standards,[29]

[24] John Harris, *'Insufficient and Very Defensive': How Nick Clegg Became the Fall Guy for Facebook's Failures*, GUARDIAN (Oct. 15, 2021, 4:56 AM), https://www.theguardian.com/politics/2021/oct/14/insufficient-very-defensive-how-nick-clegg-became-fall-guy-facebook-failures [https://perma.cc/W6ZM-LF5F] (citing Facebook's vice-president for global affairs and communications, Nick Clegg, as saying "a third of the world's population [is] on our platforms"). Despite having users all over the globe, Facebook disproportionately invests the bulk of its content moderation resources in users in the United States and Western Europe as is discussed further *infra* Section I.C.

[25] *See, e.g.*, *Transparency Reports*, META, https://transparency.fb.com/data/ [https://perma.cc/8H6K-YED2] (last visited Nov. 1, 2022).

[26] *See, e.g.*, Nina I. Brown, *Regulatory Goldilocks: Finding the Just and Right Fit for Content Moderation on Social Platforms*, 8 TEX. A&M L. REV. 451, 477 (2021) ("The major social networks have taken a mixed approach to content moderation, using both algorithms and humans to remove harmful content from their platforms. Sites such as Facebook, Twitter, and YouTube use algorithms to detect content suspected of violating their community standards.").

[27] "Machine learning is an umbrella term to describe a special subset of algorithms wherein a computer is programmed to revise the code it is using as it works, based on the results it is generating." Amy B. Cyphert, Tinker*-ing with Machine Learning: The Legality and Consequences of Online Surveillance of Students*, 20 NEV. L.J. 457, 461 (2020).

[28] *How Technology Detects Violations*, META (Jan. 19, 2022), https://transparency.fb.com/enforcement/detecting-violations/technology-detects-violations/ [https://perma.cc/K9EE-26SN] ("We remove millions of violating posts and accounts every day on the Facebook app and Instagram. Most of this happens automatically, with technology working behind the scenes to remove violating content—often before anyone sees it.").

[29] *How Meta Prioritizes Content for Review*, META (Jan. 26, 2022), https://transparency.fb.com/policies/improving/prioritizing-content-review/ [https://perma.cc/M73G-MZNZ] ("When someone posts on Facebook or Instagram, our technology checks to see if the content goes against the Facebook Community Standards and Instagram Community Guidelines. In most cases, identification is a simple matter. The post either clearly violates our policies or it doesn't.").

which prohibit certain content on the site, including cyberbullying, fraud, sexual content, hate speech, graphic content, and incitement to violence.[30]

Facebook has multiple artificial intelligence teams, and those teams build machine learning models that can perform certain tasks such as recognizing that a photo contains nudity or that text includes terms associated with hate speech.[31] Machine learning models are frequently predictive models, and Facebook's model is no exception.[32] In a typical machine learning process, an "algorithm works to identify patterns in the data it is examining, develop certain rules from those patterns (or 'learns' from them), and then uses those rules to categorize the next set of data it looks at."[33] Once the Facebook algorithm makes its prediction and flags a post as potentially violating the Community Standards, Facebook uses other artificial intelligence (what it calls its "enforcement technology") to determine whether any further action (such as removal, demotion, or review by a human) should be taken.[34] As Facebook notes, its predictive models generally become better over time, as the models "learn" from what human reviewers ultimately conclude about the predictions and then incorporate that knowledge into the next round of predictions.[35]

The company reports that its algorithms independently flag ninety percent of the content the site ultimately removes even before another user reports it[36] (though of course that still leaves up a staggering amount of content that violates the Community Standards, given the sheer volume of posts made to Facebook each day). If the algorithmic review is inconclusive, the content is flagged for a human to review it.[37] According to Facebook, the company uses three factors in determining

[30] META, *supra* note 6.

[31] *How Enforcement Technology Works*, META (Jan. 19, 2022), https://transparency.fb.com/enforcement/detecting-violations/how-enforcement-technology-works [https://perma.cc/NH5T-UBMF].

[32] *Id.* (noting that Facebook's teams "build machine learning models that can perform tasks, such as recognizing what's in a photo or understanding text").

[33] Cyphert, *supra* note 27, at 461–62.

[34] META, *supra* note 31 ("[A]n AI model predicts whether a piece of content is hate speech or violent and graphic content. A separate system—our enforcement technology—determines whether to take an action, such as deleting, demoting or sending the content to a human review team for further review.").

[35] *Id.* ("When we first build new technology for content enforcement, we train it to look for certain signals . . . . At first, a new type of technology might have low confidence about whether a piece of content violates our policies. Review teams can then make the final call, and our technology can learn from each human decision. Over time—after learning from thousands of human decisions—the technology becomes more accurate.").

[36] *Id.* ("OUR TECHNOLOGY FINDS MORE THAN 90% OF THE CONTENT WE REMOVE BEFORE ANYONE REPORTS IT." (capitalization in original)).

[37] Jeff Horwitz, *Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt*, WALL ST. J. (Sept. 13, 2021, 10:21 AM), https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353 [https://perma.cc/R7PD-GDP2] ("Sometimes the company's automated systems summarily delete or bury content suspected of rule violations without a human review. At other times, material flagged by those systems or by users is assessed by content moderators employed by outside companies."); *see also* META, *supra* note 29. Sometimes, the algorithmic

what content should be reviewed by humans: severity (the likelihood the content will lead to harm), virality (the speed with which the content is being shared), and the likelihood that the content does in fact violate the Community Standards.[38] The human review teams extend beyond Facebook employees. For example, the consulting firm Accenture has provided thousands of content moderators to the site.[39]

Facebook's content moderation has had well-documented failures (Facebook founder Mark Zuckerberg estimated that the wrong moderation call is made more than ten percent of the time, leading to more than 300,000 content moderation mistakes per day).[40] For example, the algorithms that are trained to detect nudity initially blocked pictures associated with breastfeeding and pictures of mastectomy scars meant to raise awareness for breast cancer.[41]

B. Facebook's CrossCheck Program

The process described above is Facebook's *public* account of how it does content moderation. In October of 2021, reporting in *The Wall Street Journal* revealed a Facebook program, referred to as "CrossCheck" or "Xcheck," wherein certain accounts of high-profile users (including politicians, athletes, and entertainers) are exempted from the normal processes of algorithmic content moderation, wherein "the company's automated systems summarily delete or bury content suspected of rule violations without a human review."[42] Instead, when the posts of these users are flagged as potentially violating standards, they are moderated entirely by human teams (a process sometimes referred to as "whitelisting").[43]

technology can make the moderation decision. "Other times, identification is more difficult. Perhaps the sentiment of the post is unclear, its language is particularly complex or its imagery too context-dependent. In these cases, we conduct further review using people." *Id.*

[38] META, *supra* note 29.

[39] Adam Satariano & Mike Isaac, *The Silent Partner Cleaning Up Facebook for $500 Million a Year*, N.Y. TIMES (Oct. 28, 2021), https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html [https://perma.cc/S9JX-NVB5] (noting that many of those employees "started experiencing depression, anxiety and paranoia" after working eight-hour shifts "sorting through Facebook's most noxious posts, including images, videos and messages about suicides, beheadings and sexual acts").

[40] John Koetsier, *Report: Facebook Makes 300,000 Content Moderation Mistakes Every Day*, FORBES (Jun. 9, 2020, 8:08 PM), https://www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/?sh=4f64123554d0 [https://perma.cc/XM4A-XNJ5].

[41] This example demonstrates that sometimes content moderation involves removing something that does not, in fact, violate Facebook community standards. This is known as a "take down" decision. Some content moderation mistakes involve "leave ups," where objectionable content is left up even though it should be removed (as was Prof. Putnam's experience, discussed earlier in the Article).

[42] Horwitz, *supra* note 37. In a statement to the Wall Street Journal, a Facebook spokesman said that the company was "continuing to work to phase out the practice of whitelisting," but acknowledged that criticism of the CrossCheck program was "fair." *Id.*

[43] Horwitz, *supra* note 37. The program is not a small one—documents show that the program had grown to more than 5.8 million users by 2020. *Id.* The company reportedly misled its own Oversight Board on this topic, telling the Board the CrossCheck program was used only "in a small number of decisions." *Id.*

Under the CrossCheck program, the very users whose posts have the greatest reach (and therefore perhaps the best chance to "go viral") are exempted from traditional content moderation and subjected to human content moderation, a slower process. This means that their posts might remain on the site longer, even if the posts' content clearly violates the Community Standards.[44] Indeed, this is exactly what happened under the program when a Brazilian soccer star made several posts to Facebook that included the name and nude photos of a woman who had accused him of sexual assault.[45] As *The Wall Street Journal* reported, it was content that would normally be deleted under Facebook's content moderation practices.[46] But because the soccer player was in the CrossCheck program, the "system blocked Facebook's moderators from removing [it]" until the slower CrossCheck content moderation process was finished.[47] As a result, the content remained up for more than a day and was viewed by fifty-six million users.[48]

C. Facebook's Content Moderation Outside the United States and Western Europe

In October 2021, *The Wall Street Journal* published The Facebook Files—stories based on documents that Facebook whistleblower Frances Haugen provided.[49] The documents were later shared with a larger consortium of journalists (and are now referred to as the Facebook Papers). Reporting in the wake of the Facebook Papers revealed that Facebook's content moderation practices vary widely from country to country, in part because the company spends so much less on content moderation for users who are outside of the United States and Western Europe. Haugen told reporters that, with respect to content moderation, "[w]e think it's bad in the United States. But the raw version roaming wild in most of the world doesn't have any of the things that make it kind of palatable in the United States."[50]

This disparity has been one of the revelations of the Facebook Papers and is potentially a liability for the company under several of the statutory causes of action discussed below. "Many of [the markets where Facebook struggles with content

[44] *Id.*

[45] *Id.*

[46] *Id.* ("Facebook's standard procedure for handling the posting of 'nonconsensual intimate imagery' is simple: Delete it.").

[47] *Id.*

[48] *Id.*

[49] *Id.* Facebook's Vice President of Global Affairs, Nick Clegg, issued a statement claiming that the Journal's articles included "deliberate mischaracterizations" and conferred "egregiously false motives to Facebook's leadership and employees," but also noting that it was "absolutely legitimate" for the company to be "held to account" for how it deals with issues like algorithmic amplification. Nick Clegg, *What the Wall Street Journal Got Wrong* (Sept. 18, 2021), https://about.fb.com/news/2021/09/what-the-wall-street-journal-got-wrong/ [https://perma.cc/45CU-2Y8M].

[50] Mark Scott, *Facebook Did Little to Moderate Posts in the World's Most Violent Countries*, POLITICO (Oct. 25, 2021, 7:01 AM), https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050 [https://perma.cc/BV6N-37JT].

moderation] are in economically disadvantaged parts of the world, afflicted by the kinds of ethnic tensions and political violence that are often amplified by social media."[51] Facebook's reliance on AI systems to do much of its content moderation is partially to blame for the inferior content moderation in these countries, since these algorithmic systems "don't yet understand the nuances of language."[52]

For example, in Afghanistan, where as many as five million people are monthly users of Facebook, internal documents revealed that the company "employed few local-language speakers to moderate content, resulting in less than one percent of hate speech being taken down."[53] Further, despite those nearly five million monthly users, as of December 2020, a leaked Facebook internal report on addressing hate speech in Afghanistan warned "that users can't easily report problematic content because Facebook had not translated its community standards into Pashto or Dari, the country's two official languages."[54] As is discussed below, Facebook's uneven approach to content moderation in certain countries, especially those dealing with war and ethnic tensions, may be relevant to its ultimate liability for claims arising from human rights statutes in the absence of Section 230.

D. Algorithmic Amplification at Facebook

Facebook, like all social media companies, has an incentive to optimize user engagement with the site, since that engagement directly impacts advertising revenue, which is Facebook's main income source.[55] This profit-driven model often creates perverse incentives when it comes to policing bad content on the site. As Hany Farid, a professor at the University of California, Berkeley who has collaborated with Facebook, put it, "[w]hen you're in the business of maximizing engagement, you're not interested in truth. You're not interested in harm, divisiveness, conspiracy. In fact, those are your friends."[56] Thus, Facebook and other social media companies use algorithms that are designed to help predict what

[51] Tom Simonite, *Facebook is Everywhere; Its Moderation is Nowhere Close*, WIRED (Oct. 25, 2021, 3:35 PM), https://www.wired.com/story/facebooks-global-reach-exceeds-linguistic-grasp/ [https://perma.cc/4MZC-RHFK] ("Facebook users who speak languages such as Arabic, Pashto, or Armenian are effectively second class citizens of the world's largest social network.").

[52] *Id.*

[53] Scott, *supra* note 50.

[54] Simonite, *supra* note 51.

[55] *See, e.g.*, Mike Isaac, *Facebook Nearly Doubles Its Profit and Revenue Rises 48 Percent, as Tech Booms*, N.Y. TIMES (May 5, 2021), https://www.nytimes.com/2021/04/28/business/facebook-earnings-profit.html [https://perma.cc/8MU7-KP38] (noting that "[a]dvertising revenue . . . makes up the bulk of Facebook's income"); *see also* Roddy Lindsay, Opinion, *I Designed Algorithms at Facebook. Here's How to Regulate Them*, N.Y. TIMES (Oct. 6, 2021), https://www.nytimes.com/2021/10/06/opinion/facebook-whistleblower-section-230.html [https://perma.cc/RXX8-T77G] ("[S]ocial media platforms have a fundamental economic incentive to keep users engaged.").

[56] Karen Hao, *How Facebook Got Addicted to Spreading Misinformation*, MIT TECH. REV. (Mar. 11, 2021), https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/ [https://perma.cc/8DUM-MEKD].

users will most want to see in their newsfeeds and to "amplify" those topics and posts to keep users engaged.

The company reportedly uses a secret ranking system that allegedly is based on more than 10,000 factors.[57] This system, commonly referred to as the newsfeed algorithm, helps determine what users see by deciding what is featured prominently at or near the top of their newsfeed.[58] It also makes recommendations, including recommending pages for the user to follow or other users for them to connect with. This seems harmless enough in theory, and certainly these algorithms can produce positive outcomes like fostering social connection[59] or helping to prevent users from being spammed.[60] But one problem (as Facebook's own internal research revealed)[61] is that an algorithm that is designed to optimize user engagement too heavily is going to be more likely to show users divisive content that makes them angry, since "the posts that sparked the most comments tended to be the ones that made people angry or offended them."[62]

Facebook did not always have a newsfeed algorithm that helped determine what users see. When Facebook launched its newsfeed in 2006, it was largely chronologically based, as users saw content mostly in the order it was posted.[63] In 2009, the company began using a "relatively straightforward" ranking algorithm that tried to promote "juicier" content, such as the fact of a friend's breakup.[64] The current newsfeed algorithm was the result of a series of changes that Facebook made to try to encourage what CEO Mark Zuckerberg called "meaningful social

[57] Will Oremus, *Lawmakers' Latest Idea to Fix Facebook: Regulate the Algorithm*, WASH. POST (Oct. 12, 2021, 9:00 AM), https://www.washingtonpost.com/technology/2021/10/12/congress-regulate-facebook-algorithm [https://perma.cc/3V6H-LRAR]; *see also* Will Oremus, Chris Alcantara, Jeremy B. Merrill & Artur Galocha, *How Facebook Shapes Your Feed*, WASH. POST (Oct. 26, 2021, 7:00 AM), https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works [https://perma.cc/85H9-P8VT] ("Facebook doesn't release comprehensive data on the actual proportions of posts in any given user's feed, or on Facebook as a whole.").

[58] Oremus et al., *supra* note 57 ("The top post on a Facebook user's news feed . . . is a prized position based on thousands of data points related to the user and post itself, such as the poster, reactions and comments.").

[59] *Id.* ("Feed-ranking algorithms have their benefits. At their best, they show people posts that they're likely to find interesting, surprising or valuable, and that they might not have encountered otherwise—while filtering out the noise of humdrum updates or tedious self-promotion . . . . Some researchers say they've been instrumental to some degree in fueling social movements, from the Arab Spring to Black Lives Matter.").

[60] Lindsay, *supra* note 55 ("The use of ranking algorithms by news websites for their user comment sections, traditional cesspools of spam, has been widely successful.").

[61] Oremus et al., *supra* note 57 (noting that some of the documents leaked by Haugen showed that changes made to the news feed algorithm "had the side effect of systematically promoting posts that sparked arguments and outrage").

[62] *Id.*; *see also* Lindsay, *supra* note 55 (noting that the business model wherein engagement is monetized "ensures that these feeds will continue promoting the most titillating, inflammatory content").

[63] Oremus et al., *supra* note 57.

[64] *Id.* ("Starting in 2009, a relatively straightforward ranking algorithm determined the order of stories for each user, making sure that the juicy stuff—like the news that a friend was 'no longer in a relationship'—appeared near the top.").

interaction."[65] The idea was that the algorithm would promote content that seemed to produce engagement. Commenting on a post was a sign of deeper engagement than merely liking a post, and so posts that sparked more comments were promoted.[66] Beginning in 2017, the company began to assign different weights in its ranking algorithm to different reaction buttons, and choosing the angry emoji was weighted five times as heavily as merely liking a post.[67] The result? "Facebook became an angrier, more polarizing place."[68]

These changes may become legally relevant should plaintiffs be able to sue Facebook and other social media companies for certain common law torts, as discussed below. They are especially noteworthy in light of reporting from the Facebook Papers that suggests that the company's researchers were aware of the bad impacts that changes to the newsfeed algorithm had. As *Wall Street Journal* reporter Jeff Horwitz noted, Facebook's researchers "discovered that publishers and political parties were reorienting their posts toward outrage and sensationalism. That tactic produced high levels of comments and reactions that translated into success on Facebook."[69] A team of Facebook data scientists put it bluntly in an internal memo: Facebook's new newsfeed "has had unhealthy side effects on important slices of public content, such as politics and news," which was "an increasing liability."[70] Such revelations suggest that the company was and is on notice about the impact of its recommendation algorithms, and this could be legally relevant in a post-Section 230 world.

Any recommendation algorithms used to promote posts or recommend users to each other are designed by humans, reflect human decisions, and are intentionally deployed to optimize some objective (at Facebook, often user engagement with the site). The decision to use an algorithm in the first place is a human decision. The decision to use a more opaque algorithm, one that is more of a "black box," is a human decision. The decision to define the objective of the algorithm, to choose, for example, that an algorithm will optimize user engagement with the site, is a human decision. And the decision to keep using an algorithm, even after your own researchers have flagged serious issues, is a human decision. Each of these decisions can result in liability, not for the algorithm itself but for the humans who make the decisions and the company who employs them. Therefore, Facebook and other companies cannot defend the claims we outline below by claiming that it was the algorithm (rather than the humans who designed and deployed it) that took the actions that led to the plaintiffs' harm.

[65] *Id.*

[66] *Id.* (noting that when the algorithm was changed to optimize "meaningful social interaction," it "began to give outsize weight to posts that sparked lots of comments and replies").

[67] *Id.*

[68] *Id.*

[69] Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead.*, WALL ST. J. (Sept. 15, 2021, 9:26 AM), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [https://perma.cc/5BJB-WZ2Y].

[70] *Id.*

II. CHANGE IS COMING

A. A Rare, Bipartisan Moment

Bipartisan agreement has declined in the United States for fifty years,[71] and today's leading political parties seem farther apart than ever. And yet, Republicans and Democrats both agree on the need to increase regulation of Big Tech, especially social media platforms like Facebook or Twitter. Politics can make for strange bedfellows, and the move to regulate Big Tech is no exception. Tim Wu, a Columbia Law Professor who President Biden named to the National Economic Council as a special assistant for technology and competition policy, acknowledged that the desire to regulate Big Tech has united progressives and conservatives and created some "unusual constituencies."[72] Democratic Senator and onetime presidential hopeful Amy Klobuchar supports regulating Facebook.[73] So does former Republican Senator and now Governor of Florida (and possible future presidential hopeful), Ron DeSantis.[74] Former President Trump tweeted multiple times about Section 230 prior to his ban from Twitter, exclaiming "REVOKE 230!"[75] Likewise, while he was running against then President Trump, now President Joe Biden told the editors of the *New York Times* that Section 230 should be "revoked, immediately."[76]

Of course, bipartisan agreement is not enough to create effective regulation, and the progress to date in Congress has not been especially impressive. One report described Congress' actions surrounding Big Tech regulation as an "ineffective

[71] James D. Bryan & Jordan Tama, *The Prevalence of Bipartisanship in U.S. Foreign Policy: An Analysis of Important Congressional Votes*, 59 INT'L POL. 874, 874 (2021), https://link.springer.com/content/pdf/10.1057/s41311-021-00348-7.pdf [https://perma.cc/GDL5-732Z] ("Over the past 50 years, partisan polarization has steadily increased in American politics, becoming the dominant feature of political life in the USA.").

[72] Nellie Bowles, *Fighting Big Tech Makes for Some Uncomfortable Bedfellows*, N.Y. TIMES (July 14, 2019), https://www.nytimes.com/2019/07/14/technology/big-tech-strange-bedfellows.html [https://perma.cc/N5CG-58TR].

[73] Monique Beals, *Facebook Can 'Broadly' Accept Regulators Having Access to Algorithms, Says Executive*, HILL (Oct. 10, 2021, 5:03 PM), https://thehill.com/homenews/sunday-talk-shows/576116-facebook-vp-says-broadly-the-answer-is-yes-to-regulators-accessing/ [https://perma.cc/Z2NC-SYJF]. Senator Amy Klobuchar is cited as saying, with respect to regulating Facebook, that "I believe the time for conversation is done. The time for action is now." *Id.*

[74] DeSantis signed a bill designed to "stop censorship" of Floridians by Big Tech. Press Release, Ron DeSantis, Governor of Fla., Governor Ron DeSantis Signs Bill to Stop the Censorship of Floridians by Big Tech (May 24, 2021), https://www.flgov.com/2021/05/24/governor-ron-desantis-signs-bill-to-stop-the-censorship-of-floridians-by-big-tech/ [https://perma.cc/ZB7X-LMJM].

[75] Steven Nelson, *Trump Writes 'Revoke 230!' After Twitter Masks George Floyd Tweets*, N.Y. POST (May 29, 2020, 12:46 PM), https://nypost.com/2020/05/29/trump-writes-revoke-230-after-twitter-masks-tweets/ [https://perma.cc/W3VW-83LX].

[76] Editorial Board, Opinion, *Joe Biden*, N.Y. TIMES (Jan. 17, 2020), https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html [https://perma.cc/AXZ2-PPJ7].

bricolage of finger-pointing, [and] performative hearings grilling various CEOs"[77] and noted that the debate "has been mostly unproductive and riddled with outlandish proposals."[78]

And yet, even with the dismal track record of effective action on this issue, we still argue that some form of regulation of social media companies is imminent, whether it comes from Congress or from a Supreme Court decision that changes the current scope of Section 230. Although policy makers often disagree on *why* regulation needs to happen, and often disagree on the *form* of that regulation, the existence of bipartisan agreement to "do something" coupled with rising public awareness (and outrage) of the problem makes it a safe bet to predict that some form of government regulation is coming for social media companies.[79] Although many events led to this rare bipartisan moment, the impetus to create change was crystallized in the fall of 2021 when the documents leaked by the "Facebook whistleblower" first became public. Frances Haugen worked as a data scientist at Facebook for nearly two years before leaving in early 2021. While still an employee, she copied thousands of pages of internal documents, which she ultimately shared with lawmakers, regulators, and reporters. The documents include internal Facebook research into topics like how human traffickers use the platform or how Instagram impacts teen mental health. In addition, given the 2018 legislative amendments that carve out sex trafficking claims from Section 230's protections,[80] Congress has already demonstrated its willingness to change the statute.

B. Why Section 230 Is the Most Likely Vehicle for Reform

There are a variety of forms that Big Tech regulation and reform *could* take. We present below the most widely adopted views, briefly covering ideas like antitrust enforcement and regulating Big Tech as a public utility. However, as this Section will also demonstrate, we believe that Section 230 reform is the most likely avenue of reform and has the most impact on algorithmic liability. As such, the bulk of this Section discusses why we believe Section 230 reform will occur. The looming questions are whether that change will come from Congress or the courts and what such changes might mean for a user who is seeking to hold a platform liable for algorithmic harm.

[77] Chris Riley & David Morar, *Legislative Efforts and Policy Frameworks Within the Section 230 Debate*, BROOKINGS: TECHSTREAM (Sept. 21, 2021), https://www.brookings.edu/techstream/legislative-efforts-and-policy-frameworks-within-the-section-230-debate/ [https://perma.cc/J8JE-HJ75].

[78] *Id.*

[79] Of course, it is possible that Congress may not act to amend Section 230 if it is satisfied with the outcome of the Supreme Court's decision in Gonzalez v. Google, which should be handed down by mid-2023 at the latest.

[80] *See infra* Section II.C.

1. Antitrust

Many lawmakers have introduced bills aimed at regulating Big Tech through a reimagining of antitrust laws. In June of 2021, a bipartisan group of House lawmakers introduced five bills aimed to strengthen antitrust enforcement and decrease monopolistic behavior within tech companies.[81] And in August of 2021, the FTC amended its complaint against Facebook, alleging monopolistic behavior and that the company "resorted to an illegal buy-or-bury scheme to maintain its dominance."[82] Although there may be reasons to be more skeptical about mergers and to have better laws in place to prevent monopolistic behavior, simply "breaking up" a company like Facebook is unlikely to solve the problems that arise from algorithmic recommendation systems, a point that whistleblower Frances Haugen made in her October 2021 testimony before Congress.[83] Haugen testified that breaking up Facebook under antitrust principles could have several unintended consequences, including that the advertising dollars that currently go to Facebook would likely just divert to Instagram, also owned by parent company Meta. Haugen testified that breaking up Facebook would not remove the dangerous algorithmic amplification that occurs on the site but *would* remove some of the content moderation resources.[84]

[81] Cecilia Kang, *Lawmakers, Taking Aim at Big Tech, Push Sweeping Overhaul of Antitrust*, N.Y. TIMES (June 29, 2021), https://www.nytimes.com/2021/06/11/technology/big-tech-antitrust-bills.html [https://perma.cc/22GJ-CDAQ] (noting that the bills, if passed, "would be the most ambitious update to monopoly laws in decades").

[82] Press Release, Fed. Trade Comm'n, FTC Alleges Facebook Resorted to Illegal Buy-or-Bury Scheme to Crush Competition After String of Failed Attempts to Innovate (Aug. 19, 2021), https://www.ftc.gov/news-events/press-releases/2021/08/ftc-alleges-facebook-resorted-illegal-buy-or-bury-scheme-crush [https://perma.cc/7TLZ-VDNN]. The FTC was forced to amend its complaint after a federal district court judge dismissed its original claims of monopoly against Facebook. *See* Fed. Trade Comm'n v. Facebook, Inc., 560 F.Supp.3d 1, 4–5 (D.D.C. 2021).

[83] Emily Birnbaum & Leah Nylen, *House Antitrust Leaders Meet with Facebook Whistleblower*, POLITICO (Oct. 21, 2021, 3:48 PM), https://www.politico.com/news/2021/10/21/house-antitrust-leaders-meet-with-facebook-whistleblower-516556 [https://perma.cc/T6S4-WXUX] (reporting on Haugen's congressional testimony, and noting that she said "[a] company with such frightening influence over so many people, over their deepest thoughts, feelings, and behavior needs real oversight . . . . These systems are going to continue to exist and be dangerous even if broken up").

[84] Hannah Towey, *Facebook's Week of Scandals Has Made It Easier Than Ever to Argue for Its Downfall—Here's Why the Whistleblower Still Thinks It Shouldn't Be Broken Up*, INSIDER (Oct. 5, 2021, 1:25 PM), https://www.businessinsider.com/facebook-whistleblower-testimony-frances-haugen-antitrust-instagram-break-up-2021-10 [https://perma.cc/GUM3-4FH7].

2. Public Utility

Multiple scholars,[85] a state attorney general,[86] and even one Supreme Court Justice[87] suggest that the internet in general,[88] and perhaps social media companies in particular, can and possibly should be regulated as public utilities. The basic thrust of the argument is that broadband internet and certain internet service providers like Google and Facebook are so essential to our daily lives that they have become like electricity or the telephone system, and so must be regulated by the government as those services are. The experiences many Americans had with remote work or remote school during the Covid-19 pandemic accelerated the idea that internet access is an essential good in today's economy.[89] However, labeling these companies as public utilities is not an easy endeavor and may not address crucial issues in any event. For example, although many people in parts of the world do get much of their internet news through Facebook, access to Facebook is not the same as access to the internet.

C. Section 230 Reform Is Crucial

Section 230 of the Communications Decency Act of 1996 provides that "interactive computer services" shall not be treated as "the publisher or speaker" of information provided by another person.[90] It is clear from the text of the statute that Congress was concerned at the time of its passage with protecting the

[85] *See, e.g.*, Dipayan Ghosh, *Don't Break Up Facebook—Treat It Like a Utility*, HARV. BUS. REV. (May 30, 2019), https://hbr.org/2019/05/dont-break-up-facebook-treat-it-like-a-utility [https://perma.cc/A5C3-ZUT3] (arguing that "consumer internet" like Facebook is a "natural monopoly" that should be regulated like railways and telecommunications firms).

[86] Dave Yost, Opinion, *Let's Make Google a Public Good*, N.Y. TIMES (July 7, 2021), https://www.nytimes.com/2021/07/07/opinion/google-utility-antitrust-technology.html [https://perma.cc/29KM-RLC4] (Op-ed by Ohio Attorney General urging other states to join Ohio in seeking a declaration of Google as a public utility that would have "a legal duty to act with consideration of the public interest, to provide equal access to all users and all information providers and to act without unreasonable bias against information providers").

[87] Jon Brodkin, *Clarence Thomas Blasts Section 230, Wants "Common-Carrier" Rules on Twitter*, ARS TECHNICA (Apr. 5, 2021, 3:10 PM), https://arstechnica.com/tech-policy/2021/04/clarence-thomas-blasts-section-230-wants-common-carrier-rules-on-twitter/ [https://perma.cc/ZY69-57QC] (discussing Justice Thomas's concurring opinion regarding "regulating [social media] platforms as common carriers").

[88] And now, at least with respect to the internet in general, the law has confirmed that it is an essential service. Hernán Galperin, *Infrastructure Law: High-Speed Internet Is as Essential as Water and Electricity*, CONVERSATION (Nov. 17, 2021, 8:19 AM), https://theconversation.com/infrastructure-law-high-speed-internet-is-as-essential-as-water-and-electricity-171782 [https://perma.cc/77RH-YTFR] (arguing that the crucial feature of the Infrastructure Investment and Jobs Act is its designation of the internet as "essential").

[89] David Lazarus, *The Pandemic Makes Clear It's Time to Treat the Internet as a Utility*, L.A. TIMES (Oct. 23, 2020, 5:00 AM), https://www.latimes.com/business/story/2020-10-23/coronavirus-internet-is-a-utility [https://perma.cc/5LWH-PCXC] (noting how many Americans relied on internet for work or school or shopping during the pandemic and concluding "the internet has grown into a utility, and internet access should be regulated as such").

[90] 47 U.S.C. § 230(c)(1).

development of internet services from the stifling effect of traditional tort law concepts like defamation, and it is "widely acknowledged" that the law was passed in response to a court decision holding an online bulletin board strictly liable for defamatory posts made by users.[91] It is also clear that Congress had an optimistic view of the future of discourse on the internet. The findings that begin the Section describe the internet as "an extraordinary advance in the availability of educational and informational resources to our citizens,"[92] and as something that offers "a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity."[93] Congress' optimistic view of the internet had surely dimmed by 2018, when legislators found it necessary to explicitly state that Section 230 "was never intended to provide legal protection to websites that unlawfully promote and facilitate prostitution and websites that facilitate traffickers in advertising the sale of unlawful sex acts with sex trafficking victims,"[94] and so amended Section 230 to remove the immunity for claims related to sex trafficking.

1. Courts Have Recognized Sweeping Immunity Under Section 230

Courts have consistently construed Section 230 broadly,[95] and have uniformly recognized that social media platforms are "interactive computer services"[96] and are thus immunized under Section 230 from civil liability for the content posted by their users to the site. Just a year after the passage of the Communications Decency Act, the Fourth Circuit held that AOL was an interactive computer service entitled to immunity under the Act for a negligence case brought by an AOL user.[97] The plaintiff in that case sued AOL for its delay in taking down an anonymous online bulletin board post that claimed the plaintiff was selling t-shirts that had "offensive

[91] *In re* Facebook, Inc., 625 S.W.3d 80, 89 (Tex. 2021) ("It is widely acknowledged that Section 230's liability protections were primarily a response to [a case where] a New York court held that an online bulletin board could be held strictly liable for third parties' defamatory posts.").

[92] 47 U.S.C. § 230(a)(1).

[93] *Id.* § 230(a)(3).

[94] Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No, 115-164, 132 Stat. 1253 (2018).

[95] *See, e.g.*, Force v. Facebook, Inc., 934 F.3d 53, 64 (2d Cir. 2019) ("[T]he Circuits are in general agreement that the text of Section 230(c)(1) should be construed broadly in favor of immunity."), *cert. denied*, 140 S. Ct. 2761 (2020); Michael R. Bartels, Note, *Programmed Defamation: Applying § 230 of the Communications Decency Act to Recommendation Systems*, 89 FORDHAM L. REV. 651, 658 (2020) ("Ever since § 230's enactment, courts have interpreted the statute broadly, providing nearly unlimited immunity for [interactive computer services] when third parties do harm through their conduits.").

[96] The Communications Decency Act defined "interactive computer services" as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions." 47 U.S.C. § 230(f)(2).

[97] Zeran v. Am. Online, Inc., 129 F.3d 327 (4th Cir. 1997).

and tasteless slogans" related to the Oklahoma City attack.[98] The post included the plaintiff's telephone number, and he received many phone calls, including some death threats (these calls increased after a DJ at a local Oklahoma City radio station read the AOL post on air and encouraged listeners to call the plaintiff's number.)[99] The plaintiff alleged that AOL had taken too long to remove the post and had refused to post a retraction in its place. The Fourth Circuit affirmed the dismissal of all charges, claiming they were foreclosed by Section 230.[100]

Courts have continued to define the term broadly in the years since, with cases declaring Grindr, Twitter, MySpace, and Amazon all to be interactive computer services under Section 230.

Facebook was first judicially recognized as being protected by Section 230 in a 2013 case brought against Facebook and its founder, Mark Zuckerberg. In *Klayman v. Zuckerberg*, the D.C. Circuit Court of Appeals affirmed a D.C. District Court ruling holding that Facebook is an interactive computer service because it "is a service that provides information to 'multiple users' by giving them 'computer access [ ] to a computer server,' . . . namely the servers that host its social networking website."[101] The plaintiff had brought claims of intentional assault and negligence against Facebook due to content on the site that called for violence against Jewish people.[102] Because the complaint made no allegation that Facebook had "provided, created, or developed any portion of the [allegedly harmful] content,"[103] the court concluded that the case was a relatively straightforward one.[104] Specifically, the court stated that "it is enough here to hold that a website does not create or develop content when it merely provides a neutral means by which third parties can post information of their own independent choosing online."[105]

That Section 230 immunizes Facebook from lawsuits based on the content of what an independent third-party user posts is thus clear enough, at least under Section 230 as it currently exists. But what about the decisions, often driven by algorithms, to either promote or demote certain content posted by others in users' news feeds? Are those decisions protected by Section 230 as well? Courts have thus far said yes (though not without dissenters, as discussed below), concluding that since Section 230 explicitly provides immunity for "good-faith" restrictions to user-generated content,[106] it also provides immunity for any decision to promote or demote content in a user's feed.

[98] *Id.* at 329.

[99] *Id.*

[100] *Id.* at 330–32.

[101] Klayman v. Zuckerberg, 753 F.3d 1354, 1357 (D.C. Cir. 2014) (citing 47 U.S.C. § 230(f)(2)).

[102] *Id.*

[103] *Id.*

[104] *Id.* (noting that the case "present[ed] no occasion to address the outer bounds of preemption under the Act").

[105] *Id.*

[106] 47 U.S.C. § 230(c)(2)(A).

a. Force v. Facebook

In *Force v. Facebook*, a divided Second Circuit panel affirmed a district court opinion holding that Facebook was immune under Section 230 to claims of civil liability for anti-terrorism claims stemming from certain Hamas-led attacks against American citizens in Israel.[107] The plaintiffs alleged that "Facebook unlawfully provided Hamas . . . with a communications platform that enabled those attacks."[108] The defendants moved to dismiss, and the district court agreed, holding that Section 230 foreclosed any liability against Facebook for the plaintiffs' claims. On appeal, the plaintiffs argued that Section 230 did not apply because they were not seeking to hold Facebook responsible as a *publisher* or *speaker* of Hamas's content, but rather as an involved party who "contributed to that content through its algorithms."[109]

The majority panel noted that Facebook-developed algorithms determined what content its users would see in their newsfeeds.[110] It also accepted as true for purposes of the motion to dismiss that "Facebook's algorithms directed [Hamas's content] to the personalized newsfeeds of the individuals who harmed the plaintiffs."[111] Nevertheless, the majority affirmed the district court ruling granting the motion to dismiss. The court held that Facebook's use of algorithms to impact the content that its users would see in their newsfeeds did not render it a "non-publisher" under Section 230.[112] The majority noted that "arranging and distributing third-party information inherently forms 'connections' and 'matches' among speakers, content, and viewers of content" and that such was an "essential result of publishing."[113] The majority also rejected plaintiffs' argument about Facebook's use of algorithms to curate the content its users saw in their newsfeeds, citing to earlier decisions that held that automated curating was protected under Section 230.[114] Finally, the majority rejected Plaintiff's argument that Facebook itself was an information content provider for some of the Hamas content because its algorithms had funneled the content to certain users who would be most

[107] Force v. Facebook, Inc., 934 F.3d 53, 57 (2d Cir. 2019), *cert. denied*, 140 S. Ct. 2761 (2020).

[108] *Id.*

[109] *Id.* at 62. The plaintiffs had argued that "Facebook's 'newsfeed' uses algorithms that predict and show the third-party content that is most likely to interest and engage users," and that "Facebook's advertising algorithms and 'remarketing' technology allow advertisers to target ads to its users who are likely most interested in those ads." *Id.* at 65.

[110] *Id.* at 58 (noting that "newsfeed algorithms—developed by programmers employed by Facebook—automatically analyze Facebook users' prior behavior on the Facebook website to predict and display the content that is most likely to interest and engage those particular users").

[111] *Id.* at 59.

[112] *Id.* at 66.

[113] *Id.*

[114] *Id.* at 67 (citing Marshall's Locksmith Serv. Inc. v. Google, LLC, 925 F.3d 1263, 1271 (D.C. Cir. 2019)).

receptive to the message.[115] The majority noted that Facebook did not edit the content and that its terms of service made clear that a user owned the information the user posted to the site.[116] It also held that Facebook's algorithms are content neutral because they "take the information provided by Facebook users and 'match' it to other users . . . based on objective factors applicable to any content."[117] Accordingly, it affirmed the district court decision to dismiss the claims as barred by Section 230.

Chief Judge Katzmann concurred in part with the majority panel's decision but dissented with respect to the conclusion that Facebook's algorithms deserved Section 230 publisher protection.[118] Judge Katzmann argued that Congress never intended to shield recommendation algorithms from liability[119] and noted that the plaintiffs were not bringing claims against Facebook as a publisher of content but rather seeking to hold the company liable for "its affirmative role in bringing terrorists together."[120] Judge Katzmann concluded that "[w]hen a plaintiff brings a claim that is based not on the content of the information shown, but rather on the connections Facebook's algorithms make between individuals, [Section 230] does not and should not bar relief."[121] Judge Katzmann lamented that this ruling and others like it would immunize Facebook and other social media companies from liability for harm that resulted from their algorithms and went so far as to repeatedly encourage Congress to revisit Section 230 as a result.[122] His dissent was lengthy and detailed, and included a discussion of how Facebook's algorithm pushes users toward provocative content in order to optimize site engagement.[123]

b. Gonzalez v. Google

The Supreme Court denied certiorari in *Force*. Two years later, a Ninth Circuit majority agreed with the *Force* majority, reaching a similar conclusion in *Gonzalez v. Google LLC*, holding that "a website's use of content-neutral algorithms, without

[115] *Id.* at 68 ("Plaintiffs contend that Facebook's algorithms 'develop' Hamas's content by directing such content to users who are most interested in Hamas and its terrorist activities, without those users necessarily seeking that content.").

[116] *Id.* at 69–70.

[117] *Id.* at 70.

[118] *Id.* at 76 ("I must respectfully part company with the majority on its treatment of Facebook's friend-and content-suggestion algorithms [under Section 230].").

[119] *Id.* at 77 (noting that the majority opinion "extend[ed] a provision that was designed to encourage computer service providers to shield minors from obscene material so that it now immunizes those same providers for allegedly connecting terrorists to one another").

[120] *Id.*

[121] *Id.*

[122] *Id.* ("Congress may wish to revisit the CDA to better calibrate the circumstances where such immunization is appropriate and inappropriate in light of congressional purposes."); *id.* at 84 ("[I]t therefore may be time for Congress to reconsider the scope of § 230."); *id.* at 88 ("Whether, and to what extent, Congress should allow liability for tech companies that encourage terrorism, propaganda, and extremism is a question for legislators, not judges.").

[123] *Id.* at 87.

more, does not expose it to liability for content posted by a third-party."[124] As with *Force*, the allegations in *Gonzalez* dealt with allegations of international terrorism. The *Gonzalez* plaintiffs were family members of victims of three ISIS attacks spread across the globe.[125] The plaintiffs alleged that YouTube, which is owned by Google, has "become an essential and integral part of ISIS's program of terrorism."[126] They further alleged that YouTube employed a recommendation system that recommended ISIS content to users and connected them "and that, by doing so, Google assists ISIS in spreading its message."[127]

As in *Force*, the Ninth Circuit panel in *Gonzalez* was divided, with two judges agreeing with the *Force* majority and one judge agreeing with Chief Judge Katzmann's *Force* dissent.[128] But shortly before this Article went to press in November of 2022, the Supreme Court granted cert in the *Gonzalez* case. Specifically, the Supreme Court granted cert on the question: "Does section 230(c)(1) immunize interactive computer services when they make targeted recommendations of information provided by another information content provider, or only limit the liability of interactive computer services when they engage in traditional editorial functions (such as deciding whether to display or withdraw) with regard to such information?"[129] Thus, we will likely have a Supreme Court ruling sometime in 2023, at the latest, that specifically weighs in on the question of whether Section 230 protection extends to algorithmic amplification.

It was not terribly surprising that the Supreme Court granted cert in the *Gonzalez* case, given that Justice Thomas appeared on several occasions to be inviting lower court judges to reconsider the broad and sweeping nature of the holding in cases like *Force*. Although the Supreme Court denied certiorari in *Force*, Justice Thomas, writing in a later decision to deny certiorari in another Section 230 case, suggested that courts have gone too far in their decisions to award immunity to platforms for their recommendation systems. In his written decision concurring in the decision to deny certiorari in *Malwarebytes, Inc. v. Enigma Software Group, USA, LLC*,[130] Justice Thomas opined that "[e]xtending §230 immunity beyond the natural reading of the text can have serious consequences."[131] He acknowledged

[124] Gonzalez v. Google LLC, 2 F.4th 871, 896 (9th Cir. 2021).

[125] *Id.* at 879.

[126] *Id.* at 881.

[127] *Id.*

[128] *Id.* at 895 ("Our dissenting colleague argues § 230 should not immunize Google from liability for the claims related to its algorithms, which the dissent characterizes as amplifying and contributing to ISIS's originally posted content. The dissent shares the views expressed by the partial concurrence and dissent in *Force*." (citing Force v. Facebook, Inc., 934 F.3d 53, 76–89 (2d Cir. 2019) (Katzmann, C.J., concurring in part, dissenting in part))).

[129] The question presented is available on the Supreme Court's docket in the *Gonzalez* case and at https://www.supremecourt.gov/docket/docketfiles/html/qp/21-01333qp.pdf [https://perma.cc/LP3W-4JC5].

[130] 141 S. Ct. 13 (2020) (mem.) (Thomas, J., concurring), *denying cert. to* 946 F.3d 1040 (9th Cir. 2019).

[131] *Id.* at 18.

that the Supreme Court has never ruled on Section 230[132] and concluded that, when an appropriate case arose, the Court "should consider whether the text of this increasingly important statute aligns with the current state of immunity enjoyed by Internet platforms."[133]

2. Current Section 230 Reform Proposals

Thus, the judiciary may act to limit Section 230 immunity prior to Congress even passing a bill,[134] though the backlash to the *Force* decision may actually speed Congress' actions. In October of 2020, Representatives Anna G. Eshoo of California and Tom Malinowski of New Jersey introduced the Protecting Americans from Dangerous Algorithms Act, legislation that they designed "to hold large social media platforms accountable for their algorithmic amplification of harmful, radicalizing content that leads to offline violence."[135] In her press release announcing the introduction of the bill, Congresswoman Eshoo (whose district includes much of Silicon Valley) specifically referenced the *Force* decision as a catalyst for her bill.[136]

The Protecting Americans from Dangerous Algorithms Act is not the only pending legislation that would address Section 230. There are now several proposals that would alter the scope of Section 230 immunity, especially when the content at issue is connected to acts of violence or discrimination.[137] For example, Democratic Senators Mark Warner, Mazie Hirono, and Amy Klobuchar have introduced the Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms ("Safe Tech") Act. The Safe Tech Act proposes changes to Section 230 to limit liability where social media companies "enabl[e] cyber-stalking, targeted harassment, and discrimination on their platforms."[138] In his press release

[132] *Id.* at 13 ("When Congress enacted [Section 230], most of today's major Internet platforms did not exist. And in the 24 years since, we have never interpreted this provision.").

[133] *Id.* at 14.

[134] *See, e.g.*, Mark MacCarthy, *Back to the Future for Section 230 Reform*, LAWFARE (Mar. 2, 2021, 11:54 AM), https://www.lawfareblog.com/back-future-section-230-reform [https://perma.cc/LSW2-VRE8] ("If Congress does not enact Section 230 reform, the Supreme Court could well act to 'pare back' Section 230 immunity in some way, even if not in precisely the fashion that Justice Thomas would like." But also arguing "It would be better to have Congress rethink these issues anew than to see the Supreme Court establish the new regime disguised as interpretation of a 25-year-old statutory text.")

[135] Press Release, Anna Eshoo, Cal. Congresswoman, Reps. Eshoo and Malinowski Introduce Bill to Hold Tech Platforms Liable for Algorithmic Promotion of Extremism (Oct. 20, 2020), https://eshoo.house.gov/media/press-releases/reps-eshoo-and-malinowski-introduce-bill-hold-tech-platforms-liable-algorithmic [https://perma.cc/F2MS-KKWZ].

[136] *Id.* ("The bill narrowly amends Section 230 of the Communications Decency Act to remove liability immunity for a platform if its algorithm is used to amplify or recommend content directly relevant to . . . cases involving acts of international terrorism (18 U.S.C. 2333) . . . 18 U.S.C. 2333 is implicated in several lawsuits, including [the *Force* lawsuit] alleging its algorithm connected terrorists with one another and enabled physical violence against Americans.").

[137] MacCarthy, *supra* note 134 ("A popular and bipartisan approach to Section 230 reform involves piecemeal carve-outs that focus on particularly egregious online harms and illegality.").

[138] Press Release, Mark R. Warner, U.S. Sen. From Va., Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230 (Feb. 5, 2021), https://www.warner.senate.gov/

describing the legislation, Senator Warner acknowledged that the proposal was relatively modest, and that the proposed changes would "not guarantee that platforms will be held liable in all, or even most, cases," and that, even if it were to be passed, "the current legal standards for plaintiffs still present steep obstacles."[139]

Republican lawmakers have also proposed changes to Section 230, often with a focus on mandating that the platforms be politically neutral. For example, Senator Josh Hawley has proposed the Ending Support for Internet Censorship Act, which would require that social media companies be certified by the Federal Trade Commission as politically neutral or risk losing their Section 230 immunity.[140]

In sum, there is bipartisan support for amending Section 230, though little agreement on what that amendment would look like. Federal appellate courts have recently issued conflicting rulings on the legality of state's attempts to regulate the ability of platforms to perform content moderation, and the Supreme Court will likely have to weigh in with a decision.[141] It is also possible that other appellate courts will break ranks and hold that Section 230 does not protect social media company's decisions to promote or demote content, even if it does protect the

public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230 [https://perma.cc/SV8H-WSWB].

[139] *Id.*

[140] Draft Act, Ending Support for Internet Censorship Act, https://www.hawley.senate.gov/sites/default/files/2019-06/Ending-Support-Internet-Censorship-Act-Bill-Text.pdf [https://perma.cc/F5JT-2CYS] (proposing that Section 230 immunity "shall not apply in the case of a covered company unless the company has in effect an immunity certification from the Federal Trade Commission . . . that the company does not moderate information provided by other information content providers in a manner that is biased against a political party, political candidate, or political viewpoint").

[141] *Compare* NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196 (11th Cir. 2022) (finding it substantially likely that portions of a Florida law that placed content moderation restrictions on social media companies violated the First Amendment and thus affirming in part the district court grant of a preliminary injunction), *with* NetChoice, L.L.C. v. Paxton, 49 F.4th 439 (5th Cir. 2022) (finding no constitutional violations in a Texas law that placed content moderation restrictions on social media companies and accordingly vacating a district court grant of a preliminary injunction).

hosting of that content initially.[142] What, then, will claims against social media companies look like? The next Part explores potential pathways.[143]

III. AGENCY LAW PRINCIPLES

Once Section 230 is amended, either by Congress or the courts, a floodgate of claims against Facebook and other social media companies is likely to open. These claims will almost surely allege that the companies' algorithms—the content moderation algorithms and/or the amplification algorithms, both explained above in Part I—harmed the plaintiffs in some cognizable way. Courts that analyze these claims will likely consider them within the context of already-established agency law principles. Specifically, these claims could take on two forms. First, plaintiffs could allege a cause of action for *direct liability* against social media companies like Facebook—that the company's own internal negligence – perhaps the design of its site or its own decisions regarding content moderation and/or recommendation algorithms—combined with the negligence of its agents, led to plaintiffs' harm.[144] Second, plaintiffs could base their causes of action on a *vicarious liability* theory. Under a vicarious liability theory, the principal (here, again, a social media company like Facebook) would be liable for the acts of its agents regardless of Facebook's own culpability if Facebook exerted a certain level of control over its agents and

[142] The platforms have responded to this pressure by mobilizing lobbyists and attempting to forestall some of the regulations even as the executives of the companies publicly call for enhanced government intervention. *See* Lauren Feiner, *Facebook Spent More on Lobbying than Any Other Big Tech Company in 2020*, CNBC (Jan. 22, 2021, 11:41 AM), https://www.cnbc.com/2021/01/22/facebook-spent-more-on-lobbying-than-any-other-big-tech-company-in-2020.html [https://perma.cc/A694-NHVM]. For instance, Mark Zuckerberg, the founder of Facebook, testified before Congress and urged "thoughtful reform of Section 230 of the Communications Decency Act." *Disinformation Nation—Social Media's Role in Promoting Extremism and Misinformation: Hearing Before the Subcomms. on Consumer Prot. & Com. And Commc'n & Tech. of the H. Comm. on Energy and Comm.*, 117th Cong. 7 (2021), https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-U1.pdf [https://perma.cc/TEQ6-M5KZ] (testimony of Mark Zuckerberg, Chairman & Chief Exec. Officer, Facebook, Inc.). They have also responded to the vacuum of regulations by attempting various versions of self-regulation. OVERSIGHT BOARD, https://oversightboard.com/ [https://perma.cc/HR4A-RXSU] (last visited Nov. 1, 2022) (noting that the "Oversight Board Trust and supporting independent company were formed [in October 2019], establishing the institution to provide broad oversight and management of the Oversight Board"). However, these self-governance efforts are unlikely to forestall any changes to Section 230, and so we do not examine them here.

[143] As we noted at the outset, these are not all the contemplated reforms. For instance, the creation of a new agency to regulate Big Tech has been proposed by others. *See, e.g.*, Tom Wheeler, *A Focused Federal Agency is Necessary to Oversee Big Tech*, BROOKINGS (Feb. 10, 2021), https://www.brookings.edu/research/a-focused-federal-agency-is-necessary-to-oversee-big-tech/ [https://perma.cc/5JGH-P4Z3]. This and other intriguing ideas remain outside the scope of this Article.

[144] Prior to the release of the Facebook Files, we would have argued that this would have been the harder claim to pursue against Facebook since it requires **two** negligent acts: namely a negligent act on the part of the principal (in this case Facebook) and a negligent act on the part of the agent. However, as we argue above, Haugen's revelations may show that Facebook's actions were at a minimum negligent—if not reckless—in the face of the data that Facebook's own researchers had collected and presented. *See* discussion *infra* Part I and accompanying footnotes.

those agents were acting within the scope of their employment or in a way that aligned with Facebook's objectives. The new development that plaintiffs would have to grapple with is the fact that, in cases involving claims predicated on algorithmic amplification, the *agent* could be artificial intelligence.[145]

This Section first addresses the agency principles that would establish liability for algorithmic decisions and then examines the particular claims that are most likely to be alleged against social media companies. Depending on the underlying cause of action, there are a few different pathways that can use agency principles to establish liability for the company's actions. For instance, if the underlying cause of action is a tort or a statutory violation,[146] then the principles of vicarious liability—holding an actor liable for the conduct of another—and, to a certain extent, direct liability, could be used to hold social media companies accountable.

A. General Framework

"Algorithms may soon replace employees as the leading cause of corporate misconduct."[147]

As a threshold matter, we must address the question, how does an algorithm act for purposes of legal liability? Specifically, can an algorithmic system act negligently or intentionally? Although this question is largely outside the scope of this Article (and, indeed, is large enough to form the basis for an Article of its own), it merits a brief discussion here. There are two ways courts might answer this question. First, courts might follow what we label an *analogous human* paradigm. Specifically, if the court finds that the algorithmic "conduct" in question (if undertaken by a human) would have led to liability, then liability should automatically apply.[148] Second, we could foresee a *designer assessment* paradigm that is developed whereby courts trace the negligent act to a specific human's actions—such as negligent design of the algorithm or negligent supervision of it.[149]

[145] Chiu and Lim, *supra* note 19, have argued that the deployment of machine learning algorithms makes human agency over decisions "one step removed." *Id.* at 363. While we generally agree with that premise, we do not believe that it affects our analysis. Traditional agency principles have frequently provided liability for principals not only for actions taken by their agents but also for actions taken by those far removed from the principal, in essence for sub-agents in the initial cause of action.

[146] Principles of direct and vicarious liability would both be operational in harms under either common law or statutory tort claims. If the claims against Facebook were for breaches of contract, then different agency principles would apply. A breach of contract analysis is beyond the scope of this Article.

[147] Diamantis, *supra* note 23, at 893.

[148] In that sense, we are not alone. Vladeck, *supra* note 19; *see also* Karni A. Chagal-Feferkorn, *How Can I Tell if My Algorithm Was Reasonable?*, 27 MICH. TECH. L. REV. 213, 217 (2021) ("One particularly important question for torts is whether the reasonableness analysis ought to apply when the tortfeasor is not a person, but rather a 'self-learning,' 'autonomous,' or 'artificially intelligent' system."). In fact, there is precedence for this path in the current Restatement principles (namely, that if the act, if undertaken by the principal would have been negligence, then negligence would apply. Notably, the Restatement does not make the distinction between whether the actionable conduct itself was undertaken by human or machine).

[149] Prof. Chagal-Feferkorn, discusses exactly that point in his article. *See* Chagal-Feferkorn *supra* note 148, at 249.

Ultimately, as discussed above, all algorithmic systems are the product of human decisions and reflect those decisions. Thus, we believe that there are factual scenarios which could give rise to social media company liability under either theory.[150] As such, the claims we discuss below focus on how and why a corporation like Facebook should be held liable for the outputs of its algorithms.

How might an algorithm be liable under corporate law doctrine? Perhaps surprisingly, it might not be difficult to conceptualize.[151] Legal precedent from tort law already exists, and scholars have begun mapping algorithmic liability onto that existing precedent,[152] a process we continue with this Article. Corporate law jurisprudence in the United States has a number of different principles that may apply—here, we discuss them within the context of agency law principles.[153]

The bedrock of agency law is the idea that the law has found some relationship between two parties that requires us to treat the actions of the agent as if it were undertaken by the principal.[154] This, in turn, is based on notions of consent and convenience—namely, both parties consenting to the arrangement in question, usually for the convenience of the principal. Or, as the Restatement (Third) of Agency puts it:

> Agency is the fiduciary relationship that arises when one person ("a principal") manifests assent to another person (an "agent") that the agent shall act on the principal's behalf and subject to the

[150] Nonetheless, we also recognize that we do not know what we do not know. Specifically, given the rapid evolution of technology, at least one of us foresees a situation where the algorithm might be found negligent despite all due care being exercised by human designers.

[151] Indeed, the concept of algorithmic (or, in the words of one author "computational" agency) has already been discussed. *See, e.g.*, Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203, 207 (2015) (introducing the concept of computational agency and discussing the role of algorithms as manipulative gatekeepers); *see also* Diamantis, *supra* note 23, at 893.

[152] *See, e.g.*, Pinchas Huberman, *A Theory of Vicarious Liability for Autonomous-Machine-Caused Harm*, 58 OSGOODE HALL L.J. 233 (2021) ("[T]he doctrinal form of vicarious liability is a promising strategy to ground tort liability for autonomous-machine-caused harm.").

[153] Although it is outside the scope of this Article, there may also be a pathway to liability for social media companies like Facebook based on other principles such as (1) products liability concepts, (including defects in the design and implementation (or manufacturing) of algorithms), Vladeck *supra* note 19, at 142; and, (2) corporate personhood and unpacking the role of corporate personhood within the context of AI, Alicia Lai, *Artificial Intelligence, LLC: Corporate Personhood as Tort Reform*, 2021 MICH. ST. L. REV. 597; Nadia Banteka, *Artificially Intelligent Persons*, 58 HOUS. L. REV. 537 (2021) (assessing the use of legal personhood and its ability to apply to AI).

[154] An early commentary on the subject discusses (and challenges) the notion within the context of the Latin term *qui facit per alium facit per se*, "he who acts through another does the act himself." F.E. Dowrick, *The Relationship of Principal and Agent*, 17 MOD. L. REV. 25, 24 (1954). In his seminal article on the subject, Dowrick goes on to argue that most scholars were incorrectly looking at agency relationships though a contractual lens, (either implied or express contract), however, by examining English law (from which American common law derives), he shows that the relationship is much more expansive than that. As Dowrick points out, "it is true that in almost all cases a contract accompanies an agency, but there may be a complete agency without a contract." *Id.* at 26. As discussed above, the modern definition of agency supports this position.

> principal's control, and the agent manifests assent or otherwise consents to act.[155]

Although agency principles can be used in a variety of contexts, they are frequently seen (and overlap with) corporate law jurisprudence when various actors, working on behalf of the corporation, commit acts that are then used to ascribe liability to the corporation itself.[156] Given the non-human nature of the corporation, this must be so.[157] Since a corporation can only act through intermediaries, this agency model of liability flows from the fact that these intermediaries commit acts as agents of the corporation—acts that are, in turn, attributed to the corporate entity.[158] Until recently, the intermediaries whose actions the law was concerned with were easy to identify; they were humans (usually, but not always, employees of the corporation) who were vested with authority to act or whose actions were in some way controlled by the corporation.[159] As such, the law ascribed their actions to the principal, namely the corporation itself.[160] However, in recent years, scholars have become more concerned with the ever-growing reality that because of the advances in artificial intelligence, it is often the actions of non-human intermediaries that lead to the negligent act in question. In either scenario, once the agency relationship has been established and the underlying cause of action has been determined, the focus then returns to principal. Specifically, the next step in the analysis would examine whether the principal can be held liable under theories of either vicarious or direct liability. If either theory is proven, the result will be the same: the principal will be held liable for the actions of its agent.

1. Direct Liability

Under certain circumstances, a company is directly liable for the acts of its agent. For instance, under the Restatement of the Law (Third) on Agency, a principal is directly liable for an agent's conduct when (1) the agent acts with actual authority,[161] or the principal in some way ratifies the agent's conduct, and (2) the conduct is tortious. For instance, if a principal explicitly directs her agent to commit a tort (e.g., where a nefarious businessman directly authorizes a friend to injure another) then both the principal and the agent would be held responsible under a

[155] RESTATEMENT (THIRD) OF AGENCY § 1.01 (AM. L. INST. 2006). This definition almost mirrors the previous restatement's definition regarding the agency relationship, *see* RESTATEMENT (SECOND) OF AGENCY (AM. L. INST. 1958), demonstrating that this principle has long been established.

[156] *See, e.g.*, Meyer v. Holley, 537 U.S. 280 (2003) (discussing traditional agency principles within the context of corporate liability).

[157] *Id.* at 286 (first citing RESTATEMENT (SECOND) OF AGENCY § 219(1) (AM. L. INST. 1958); then citing 3A W. FLETCHER, CYCLOPEDIA OF THE LAW OF PRIVATE CORPORATIONS § 1137 (rev. ed. 1991–1994); and then citing 10 *id.*, § 4877 (rev. ed. 1997–2001)).

[158] *Id.*

[159] *Id.*

[160] *Id.*

[161] RESTATEMENT (THIRD) OF AGENCY § 7.04 (AM. L. INST. 2006).

direct liability theory.[162] Moreover, under the principle of direct liability, the principal could still be liable for the acts of his agent—even if the agent was not acting with explicit instructions—if the tortious conduct occurred as a result of the agent's actual authority.[163] Similarly, if the agent, at the time that the tort occurred, had no authority to act but the principal later ratifies her actions, then the law will treat the agent as if she had actual authority at the time of the tort and, as such, will ascribe liability to the principal as well.[164]

Thus, to establish direct liability for a company like Facebook based on the actions of its agent, plaintiffs would need to show that the agent (here, either a human or an algorithm) acted with actual authority from Facebook, or that the tortious acts were later ratified by Facebook. The Restatement's commentary regarding actual authority is also instructive. According to the Restatement: "When an agent acts with actual authority, the agent reasonably believes, in accordance with the manifestations of the principal, that the principal wishes the agent to act."[165] Although there has yet to be a case that establishes these parameters within the context of a social media algorithm, evidence of actual authority might be established in several ways.

For example, evidence that the algorithm was acting within the scope of its original design (the objective of the algorithm could be discerned through depositions during discovery) might suggest actual authority. For example, if Facebook's recommendation algorithm was designed to optimize user engagement more heavily than other objectives, and Facebook's own researchers warned that it was doing so at the cost of increasing anger and division, and Facebook still persisted with the algorithm, plaintiffs have a potentially powerful point to make about actual authority.

In addition, courts might look to whether the results and/or outputs of the algorithm were intended by the humans that designed and deployed it. If the results or outputs were unintended consequences, were they still reasonably foreseeable to the company? Again, as above, the allegation would not be that Facebook intended to sow division or connect people to commit violent acts. Rather, it would be

[162] *Id.*

[163] *Id.* at cmt. b. ("If an agent's action within the scope of the agent's actual authority harms a third party, the principal is subject to liability if the agent's conduct is tortious.") When teaching this concept in class, Jena uses the following scenario to illustrate: A nightclub owner is taken by surprise when a deluge of would-be patrons descends upon his venue. He asks his cousin "Bowser the Bruiser" to help regulate the line for him as a favor. Specifically, he says "do whatever you need to, but don't let more than thirty people in every hour." Bowser, eager to practice his newly found MMA skills, gets excited when a few college boys try and jump the line. As a result of Bowser's actions, one of the students ends up in the hospital. In that instance, both Bowser and the nightclub owner would be liable for the tort—even though the nightclub owner didn't explicitly authorize Bowser to use force. *See also* Soc'y of the Holy Transfiguration Monastery, Inc. v. Gregory 689 F.3d 29, 57 (1st Cir. 2012) (denying a principal's defense against liability for his agent's tortious conduct because the agent was acting with the actual authority of the principal).

[164] RESTATEMENT (THIRD) OF AGENCY § 7.04 (AM. L. INST. 2004).

[165] *Id.* at cmt. b.

enough, at least at the motion to dismiss stage, to argue that Facebook intended to increase user engagement and knew that doing so would also likely promote harmful content.[166]

For the case of ratification, the cause of action might be easier to prove. Here, plaintiffs would simply have to show that the principal knew of the outcomes that the algorithm was causing and continued to deploy that algorithm knowing its harmful results. Much of the reporting regarding the Facebook Papers has revealed internal research from Facebook's own data scientists and integrity teams showing that Facebook employees were increasingly alarmed at the impact of changes to the recommendation algorithm.[167] These documents, and others that might be revealed in discovery, would be critical in analyzing whether Facebook's actions amounted to ratification.

Key to this analysis (and in direct contrast to the discussion below) is that the plaintiffs in this case *need not* show negligence on the part of the principal itself. It would be sufficient to merely show that the agent's tortious conduct was done with actual authority (or ratification) on the part of the principal.

However, a principal can also be liable for the actions of its agent if the principal was "negligent in selecting, training, retaining, supervising, or otherwise controlling the agent."[168] Again, the revelations in the Facebook Papers might be relevant: theoretically a plaintiff can use information like this to argue that Facebook's actions in deploying its algorithms amounted to negligent supervision, particularly given how aware the company seemed to be regarding the effects of its code in the "rest of the world."[169] For instance, we could envision a scenario where discovery produces evidence that Facebook did not consider the impacts that its algorithms were having on the underlying cause of action, even when they were on notice that the algorithm was producing harmful consequences.[170]

[166] Of course, as is discussed *infra* Section III.B.3, plaintiffs would still face an uphill battle in terms of causation with such claims.

[167] *See supra* Section I.C (discussing the WSJ article's reporting on Facebook data scientists and the company's "increasing liability" for its recommendation algorithm).

[168] RESTATEMENT (THIRD) OF AGENCY § 7.05 (AM. L. INST. 2006).

[169] We could envision a scenario where a court takes into consideration the relatively little amount of manpower devoted to monitoring overseas hate speech that could be directly traceable to tortious outcomes. *See, e.g.*, *infra* Section III.B (our international hypo) and accompanying discussion.

[170] It seems as if plaintiffs are already attempting to use these theories to ascribe liability to Facebook. *See, e.g.*, Complaint, Underwood v. Meta Platforms, Inc., No. 39130.001, at ¶ 5 (Cal. Super. Ct. Jan. 5, 2022), https://www.cohenmilstein.com/sites/default/files/Complaint%20-%20January%206%2C%202022.pdf [https://perma.cc/R33Z-S67M] ("The shooting was not a random act of violence. It was the culmination of an extremist plot hatched and planned on Facebook by two men who Meta connected through Facebook's groups infrastructure and its use of algorithms designed and intended to increase user engagement and, correspondingly, Meta's profits."); *see also* Press Release, CohenMilstein, *Underwood v. Meta Platforms, Inc.* (Facebook) (2022), https://www.cohenmilstein.com/case-study/underwood-v-meta-platforms-inc-facebook [https://perma.cc/96BR-7Q44] (discussing the lawsuit the firm filed against Facebook).

2. Vicarious Liability

Vicarious liability claims seek to hold one person (the principal) liable for the actions of another "person" (the agent) based on the agent's tortious conduct.[171] A key distinction between torts based on vicarious liability and torts based on direct liability is the principal's own conduct *vis-à-vis* the underlying tort. As discussed above, in cases based on direct liability plaintiffs would have to show an additional, independent act on the part of the principal that links their actions to those of the tort. For instance, in the case of a supervisor who negligently supervises an employee, the two acts involved are (1) the tortious conduct of the employee and (2) the negligent supervision of the employer which prevented them from catching the employee's mistake.[172] In contrast, with vicarious liability the principal frequently has no direct connection to the underlying tort. Rather, the focus for establishing vicarious liability is centered on control: was the agent subject to such a significant amount of the principal's control that the agent would be deemed an employee-agent[173] of the principal? If so, then the principal can be liable (even if they were unaware of the agent's tortious actions), but only if the agent was acting within the scope of employment.[174]

Here again, the Restatement is instructive: "(2) an employee acts within the scope of employment when performing work assigned by the employer."[175] Thus, in order for a corporation to be vicariously liable for the conduct of an algorithmic system, that system would have to be employed by the corporation and also

[171] The tortious action in question most frequently sounds in claims based on negligence. However, principals have also been held liable for the intentional torts of their agents. *See e.g.*, Manning v. Grimsley, 643 F. 2d 20 (1st Cir. 1981) (discussing liability of the principal for the intentional tort of battery, based on the agent's conduct as a pitcher for the Baltimore Orioles).

[172] Similarly, where a principal explicitly authorizes an agent to act, the principal is in essence taking ownership over the agent's underlying conduct.

[173] Although calling an algorithm an employee might seem a tortured use of the term, legal scholars are already discussing that very theory. *See e.g.*, Brandon W. Jackson, *Artificial Intelligence and the Fog of Innovation: A Deep-Dive on Governance and the Liability of Autonomous Systems*, 35 SANTA CLARA HIGH TECH. L.J. 35, 57 (2019) ("Similarly, an AI system could be treated as an employee and the owner as an employer."); Huberman, *supra* note 152, at 255–56 (noting the pros and cons of treating AI systems as "employees" for vicarious liability purposes). It also helps to keep in mind that the Restatement language predates the conception of AI in any meaningful way. Another way to consider the issue is to examine whether Facebook, in fact, *employed* the algorithm. To that end, Merriam's definition is instructive: "employed. Transitive verb. 1(a) to make use of (someone or something inactive). B: to use (something such as time)." Merriam-Webster's Collegiate Dictionary (11th ed. 2008).

[174] RESTATEMENT (THIRD) OF AGENCY § 7.03 (AM. L. INST. 2006) ("A principal is subject to vicarious liability to a third party harmed by an agent's conduct when . . . the agent is an employee who commits a tort while acting within the scope of employment.").

[175] *Id.* § 7.07 ; *see also id.* ("(b) as stated in § 7.08, the agent commits a tort when acting with apparent authority in dealing with a third party on or purportedly on behalf of the principal.") As such, there is also the possibility that courts find that the algorithm acted with apparent authority, however that is not an issue we address here.

performing the work that it was assigned (or in this case, designed) to do.[176] This particular theory of liability might also prove useful for plaintiffs who are attempting to bring claims against social media companies because the evidentiary hurdles needed to establish vicarious liability in this instance would seem lower than what was needed to establish direct liability.[177] In short, all that would seem to be needed is evidence showing that (1) these companies exerted a significant amount of control over their algorithms and (2) the algorithm acted as designed.[178]

Defendants might try to use the fact that the machine learning algorithm is sometimes quite opaque—a "black box"[179]—as a defense against claims of vicarious liability. However, such arguments should not be convincing to courts. First, it would encourage companies to use especially opaque algorithms, something that is already being done[180] but would certainly increase if companies perceive their liability is lowered. Second, it would lead to a potential landscape where plaintiff recovery could never happen since employing a machine learning algorithm would arguably act as a bar against all of these types of claims. Third, it would ignore the reality discussed in Section I.D: that all algorithms, even machine learning ones, reflect human decisions and human inputs, and thus liability for their outputs should lie with the humans who make those decisions, at least when those outputs are reasonably foreseeable.

As such, if plaintiffs successfully establish the agency principles at work[181] and can prove the underlying causes of action (which, after a brief discussion of other potential sources of authority, we turn to next), then liability will likely follow.

B. Potential Authority from Other Areas of the Law

One final threshold note: although the hypothetical claims we discuss below would represent cases of first impression since they would represent litigation in a post-Section 230 world, courts looking for persuasive authority will nonetheless have some case law upon which to draw. For instance, courts could look to the body of law that has developed around autonomous vehicles, which also rely on

[176] *Id.* Certainly, it would seem difficult for Facebook (or other social media companies) to argue that the algorithm wasn't "performing work assigned" since that is exactly what the algorithm was designed to do.

[177] For an insightful discussion on the underlying theories for applying a vicarious liability framework to AI, see Diamantis, *supra* note 23, at 926.

[178] The law states "an employee is an agent whose principal controls or has the right to control the manner and means of the agent's performance of work." RESTATEMENT (THIRD) OF AGENCY § 7.07 (AM. L. INST. 2006).

[179] Ben Smith, *A Former Facebook Executive Pushes to Open Social Media's 'Black Boxes,'* N.Y. TIMES (Jan 2, 2022), https://www.nytimes.com/2022/01/02/business/media/crowdtangle-facebook-brandon-silverman.html [https://perma.cc/YCF4-D8C2].

[180] *Id.*

[181] *See* Chagal-Feferkorn, *supra* note 148, at 249 (noting that analyzing an algorithm's reasonableness standard could lead to liability on the part of the manufacturer's algorithm in a way that would be "similar to analyzing the behavior of an employee when determining whether their employer is vicariously liable for their actions").

algorithms.[182] In addition, courts could look at cases surrounding the internet of things.[183] Further, some scholars have already advocated for strict liability regimes for algorithms.[184] Each of these can provide a helpful framework for courts to assess the liability of social media companies for their algorithmic acts. Nonetheless, we believe agency principles may provide the most useful framework.

C. Particular Types of Claims

The specific claims that victims may bring against social media companies will generally fit into two broad categories: (1) specific statutory claims, and (2) various types of common law tort claims. Each of these will be discussed in turn.

There are a number of statutory claims that people have already alleged, unsuccessfully, against Facebook (and other companies) that could be resuscitated in this new liability regime. The most likely statutory claims, given the claims brought in the past, are liability for terrorist activities (based on a federal terrorism statute)[185] and liability under human trafficking laws (indeed, as was discussed above, Section 230 was amended in 2018 to exclude sex trafficking claims from the liability shield).[186] The most likely tort causes of actions include defamation and negligent infliction of emotional distress.[187]

[182] *See, e.g.*, Mark A. Chinen, *The Co-Evolution of Autonomous Machines and Legal Responsibility*, 20 VA. J.L. & TECH. 338, 363 (2016) (discussing the current lack of legal responsibility for designers of autonomous machines).

[183] Rebecca Crootof, *The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference*, 69 DUKE L.J. 583 (2019) (discussing a civil liability framework within the context of the internet of things).

[184] Vladeck argues for a strict liability regime. Vladeck *supra* note 19, at 147.

[185] 18 USC § 2333.

[186] Although, on its face, the 2018 amendment only excludes *sex* trafficking claims from Section 230's liability shield, the Texas Supreme Court recently applied this exclusion from liability to *all* trafficking claims. For a discussion of the Court's rationale and the implications of this, see *infra* Section III.C.1. For a discussion of trafficking within the context of other coercive labor practices, see Jena Martin, *Guest Blog: ULC's work on Coercive Labor Practices in Supply Chain, Part 1*, BUS. L. PROF BLOG (Aug. 16, 2020), https://lawprofessors.typepad.com/business_law/2020/08/since-1892-the-uniform-law-commission-has-deeply-affected-the-practice-of-law-especially-business-law-uniform-acts-like.html [https://perma.cc/E77Z-TJDC].

[187] For claims that amount to egregious human rights violations foreign plaintiffs may be able to avail themselves of the Alien Tort Statute (the ATS). Specifically, the ATS, a jurisdictional statute, has been used in the past to allow foreign nationals to bring claims against U.S. corporations in federal courts. 28 USC § 1350. Currently, however, this is an uphill climb. Recently, the Supreme Court has been very active in ATS jurisprudence. For instance, it has established that a foreign plaintiff cannot bring a claim against a foreign defendant for claims that occur on foreign soil. *See* Kiobel v. Royal Dutch Petroleum, 569 U.S. 108 (2013). Subsequently, the Court held that *no* cases could be brought against foreign corporations. Jesner v. Arab Bank, PLC, 138 S. Ct. 1386 (2018). Most recently, the Court in *Nestle v. Doe*, 141 S. Ct. 1931 (2021), held that the standard for overcoming the extraterritoriality presumption in *Kiobel* was not met when U.S. corporate officers made decisions that were tied to a "generic" corporate function. However, *Nestle* left open the question of whether a U.S. corporation's actions needed to have directly caused the harm in question or whether aiding and abetting was enough. As such, to successfully plead an ATS claim against Facebook under the Court's jurisprudence in *Nestle*, the plaintiffs, at a minimum, would have to show that officers in the American headquarters made specific decisions about the "rest of the world" content moderation resources knowing the potentially

To provide a more concrete analysis of how ending social media sites' protection under Section 230 might impact these claims, we provide two hypotheticals—both modeled after recent headlines.[188] One, discussed in this Section, implicates statutory violations[189] from international incidents (such as terrorism and human trafficking). The other relates to incidents here in the U.S. that could trigger common law tort claims.

For the first hypothetical, imagine that a militant group in Qumar[190] (a country in the Global South) uses Facebook to flood messages of hatred and violence against a particular social-ethnic group within that country. Posing as regular citizens on the site, the militant group (who has been designated by the U.S. as a foreign terrorist organization) orchestrates a campaign of genocide and human trafficking against the ethnic minority. Over the course of several years, operatives within the group routinely make posts regarding the ethnic minority, calling them "godless" and "not fit to live." The posts exhort other users to engage in a merciless campaign of killing, sex trafficking,[191] and forced migration. International human rights agencies term the campaign coordinated genocide. The users, located in Facebook's "rest of the world," engage in the conduct on Facebook's platform with little human oversight. Rather, because of Facebook's algorithmic amplification, these posts routinely go viral, with the most violent and vitriolic posts seemingly gaining the most attention. How might a social media company like Facebook be held liable for the harm that ensues in such a hypothetical?

disastrous consequences from a human rights standpoint. Plaintiffs who have been the victim of human rights violations could still bring their cases; however, unless they were able to avail themselves of the personal jurisdiction rules in federal court, they would be limited to state court venues. For a discussion of ATS litigation in the United States, see Rachel Chambers & Jena Martin, Potential Paths Forward After the Demise of the Alien Tort Statute, *in* Civil Remedies and Human Rights in Flux (Ekaterina Aristova & Uglesa Grusic eds. forthcoming 2022).

[188] The first hypothetical is patterned after genocide in Myanmar that occurred after viral violent posts on social media. *See* Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar's Military*, N.Y. Times (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html [https://perma.cc/C84D-X7GL]. The second hypothetical is based on reporting from *The Washington Post* discussing a QAnon conspiracy theory about an alleged child sex trafficking ring. *See* Jessica Contrera, *A Qanon Con: How the Viral Wayfair Sex Trafficking Lie Hurt Real Kids*, Wash. Post (Dec. 16, 2021), https://www.washingtonpost.com/dc-md-va/interactive/2021/wayfair-qanon-sex-trafficking-conspiracy/ [https://perma.cc/4QLR-PUKC].

[189] One advantage for a plaintiff in bringing causes of actions under statute is that it eschews a court's need to wrestle with issue such as whether an *algorithm* is acting negligently or intentionally (issues we discuss *infra* Section III.A).

[190] To be clear, Qumar is a fictitious country. However, we were not the ones that first came up with the name. *See* The West Wing: The Women of Qumar (NBC television broadcast Nov. 28, 2001).

[191] While we present the sex trafficking claims within the context of an international incident, these claims implicate both international and national law (which often share similar structures). For a discussion of the comparison of the two, see Women's Env't & Dev. Org., Trafficking of Women: U.S. Policy and International Law (2005), https://www.wedo.org/wp-content/uploads/trafficking.pdf [https://perma.cc/7CYD-76ZA].

1. Liability Under Trafficking Laws

Another potential cause of action that plaintiffs have already alleged in connection with social media conduct are claims related to trafficking. As noted above, *sex* trafficking claims have been exempted from Section 230's liability shield since 2018. Although this is a relatively recent change, and courts have not had the opportunity to issue many opinions related to trafficking claims brought against social media companies, early opinions are a useful guide for predicting how courts may respond to claims when Section 230 is amended, and also for predicting the types of claims plaintiffs will bring and hurdles they will face. Further, the way the courts address issues of causation when ruling on these claims against social media companies will be an important proverbial canary in the coal mine and provide early glimpses of how other claims may fare if they too are removed from Section 230's liability shield.

For instance, in 2021, the Texas Supreme Court cited Section 230 to dismiss common law claims of negligence and products liability against Facebook but *declined* to dismiss a state law civil claim for the role Facebook's algorithms played in connection with the sex trafficking of several victims.[192] Tex. Civ. Prac. & Rem. Code § 98.002[193] provides for civil liability for any person "who intentionally or knowingly benefits from participating in a venture that traffics another person."[194] The court did not take a specific position on whether the defendant's actions as alleged at the writ of mandamus stage did in fact amount to a violation of the state trafficking statute.[195] Rather, the court merely held that "the statutory claim for knowingly or intentionally benefiting from participation in a human trafficking venture is not barred by Section 230 and may proceed to further litigation."[196]

Tellingly, the Texas Supreme Court did not interpret the 2018 amendment to Section 230 as limited to only federal *sex* trafficking claims or state criminal trafficking claims, as the language of that amendment might suggest.[197] Rather, the Court agreed with the plaintiffs that the amendment "announces a rule of construction applicable to section 230."[198] Facebook had argued that the amendment only carved out from Section 230's liability shield "a civil action under 18 U.S.C § 1595 and certain state criminal prosecutions but not civil human

[192] *Id.*

[193] TEX. CIV. PRAC. & REM. CODE ANN. § 98.002 (West 2022).

[194] *Id.*

[195] *In re* Facebook, Inc., 625 S.W.3d 80, 83 (Tex. 2021).

[196] *Id.*

[197] The specific changes to Section 230 brought as a result of the 2018 amendment were to carve out from the liability shield: "(A) any claim in a civil action brought under section 1595 of title 18, United States Code, if the conduct underlying the claim constitutes a violation of section 1591 of that title; (B) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 1591 of title 18, United States Code; or (C) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 2421A of title 18, United States Code, and promotion or facilitation of prostitution is illegal in the jurisdiction where the defendant's promotion or facilitation of prostitution was targeted."

[198] *In re Facebook*, 625 S.W.3d at 100.

trafficking claims under state statutes,"[199] such as the one at issue in the case. The Texas Supreme Court disagreed, ultimately holding that the amendment, rather than narrowly carving out certain exceptions, instead "reflect[ed] Congress's judgment that such claims were never barred by section 230 in the first place."[200]

The plaintiffs in the Texas case alleged that Facebook harmed them by "creating a breeding ground for sex traffickers to stalk and entrap survivors," including through its advertising policies.[201] Further, the complaint alleged that Facebook violated the state trafficking statute through such "acts and omissions" as "knowingly facilitating the sex trafficking of [Plaintiffs]."[202] Certain of these allegations, including that Facebook created conditions that allowed sex traffickers to easily connect to victims, could be the result of Facebook's decisions to engage in algorithmic amplification, which promoted certain posts to certain users. As such, should similar allegations be proven by plaintiffs, then the statute—coupled with the agency principles discussed above—could ultimately lead to liability for Facebook under the Texas statute.

The Supreme Court declined to hear the plaintiffs' appeal of the Texas Supreme Court decision.[203] Thus, the Supreme Court will not weigh in (at least not yet) on whether the Texas Supreme Court was correct in expanding the 2018 amendments to exclude *all* trafficking claims from Section 230's liability shield, rather than just federal sex trafficking claims or state criminal sex trafficking claims.[204] Nonetheless, the fact that a state supreme court has allowed a state trafficking claim against Facebook to proceed to the next stages of litigation, and the fact that the Supreme Court declined to grant cert, is a powerful predictor of how courts in a post-Section 230 world will treat the remaining claims this Article examines, and underscores the urgency of this analysis.

For instance, if the language used in *In re Facebook* were to stand,[205] plaintiffs might *currently*[206] be able to bring claims under federal trafficking laws like the Trafficking Victims Protection Reauthorization Act (TVPRA). Under that Act,

[199] *Id.* at 99.

[200] *Id.*

[201] *Id.* at 85 (quoting from underlying complaints).

[202] *Id.* at 96 (quoting from underlying complaints).

[203] *See* docket for Jane Doe, Facebook, Inc., No. 21-459, https://www.supremecourt.gov/search.aspx?filename=/docket/docketfiles/html/public/21-459.html [https://perma.cc/7CK8-HUYK] (last visited Nov. 1, 2022).

[204] The language discussed in *In re Facebook* is modeled after a federal statute, but it is not the Trafficking Victims Protection Reauthorization Act. It is, as the court mentions, based on a federal statute that makes the "sex trafficking of children . . . by force, fraud, or coercion" a crime. *In re Facebook*, 625 S.W.3d at 23 (citing 18 USC § 1591). Coincidentally, the 2018 amendments also specifically reference 18 USC § 1591.

[205] As mentioned earlier, despite the plain language of the 2018 amendments, the court's language in the case is expansive. It notes that "civil liability may be imposed on websites that violate state and federal *human*-trafficking laws." *In re Facebook*, 625 S.W.3d at 83 (emphasis added).

[206] The potential confusion in the Texas Supreme Court's use of trafficking is also an example of how additional action from Congress could provide greater clarity on the reach of Section 230. For our additional recommendations, see *infra* Section IV.A.

plaintiffs can file civil actions against corporate defendants who engaged in forced labor, involuntary servitude, or sex trafficking.[207] As one of us has noted (in another co-authored piece):

> [The TVPRA] allows victims to sue perpetrators for monetary damages—both compensatory and punitive—if they establish by a preponderance of the evidence that the perpetrators benefited from participation in a venture that they knew or should have known engaged in forced labour or trafficking.[208]

Algorithm as Agent for TVPRA

Returning to our hypothetical, plaintiffs could allege that Facebook's recommendation algorithm used detailed information that the algorithm collected on Qumar's ethnic minority to connect predators to people within that ethnic minority group. Further, plaintiffs could allege that the recommendation algorithm suggested that people join groups on the site that include posts with dehumanizing language about the ethnic minority, and that such experiences radicalized them.

Hypothetical Facebook Liability

The acts that plaintiffs would need to show to lead to liability on the part of Facebook for trafficking have already been laid out by the Plaintiffs in *In re Facebook*, and include allegations like "creating a breeding ground for sex traffickers to stalk and entrap survivors."[209] Further, claims predicated on algorithmic amplification might include allegations of acts like: negligent supervision on the part of Facebook's algorithmic designers, especially if the recommendation algorithm connected potential victims to traffickers; information showing that Facebook in essence "ratified" the algorithmic conduct, which might come in the form of being on notice that the recommendation algorithm was connecting potential victims to traffickers and choosing to continue to use it regardless; or evidence showing the extensive control that Facebook had over its technology.

In short, if the plaintiffs in our hypothetical can show that Facebook's algorithms amplified content that encouraged (and indeed caused) its users to engage in human trafficking as part of the terror campaign in Qumar, there is a strong likelihood that Facebook could be held liable, especially pending the outcome of any Supreme Court ruling on the *Google v. Gonzalez* appeal.

2. The Anti-Terrorism Act (ATA)

Plaintiffs have brought claims against Facebook for terrorist activity (as in the *Force* and *Gonzalez* cases, discussed above), but those claims have thus far been

[207] Trafficking Victims Protection Reauthorization Act of 2003, Pub L No 108-193, § 4(a)(4)(A), 117 Stat 2875, 2878 (2003).

[208] CHAMBERS & MARTIN, *supra* note 187, at 359.

[209] *In re Facebook*, 625 S.W.3d at 96.

dismissed because of Section 230's liability shield. If that shield is altered or removed, perhaps by the Supreme Court when it hears the *Gonzalez* appeal, plaintiffs may bring claims under the Anti-Terrorism Act (ATA or "Act"). In 1990, Congress passed the ATA, which allowed victims to bring civil actions for harms related to terrorist activities. The statute defines international terrorism as "violent acts or acts dangerous to life"[210] that occur outside U.S. borders and that are specifically designed to—in the words of one commentator—"intimidate or coerce populations or public policy."[211] Under the ATA, victims of international terrorism can bring a lawsuit in U.S. federal court to receive redress. The statute has expanded since its original enactment.[212] In its current form, the law now extends liability to "any person" who knowingly conspires or "aids and abets" any act of international terrorism.[213] As such, the statute would provide for civil recovery against a social media company by a plaintiff who can prove: (1) an act of international terrorism;[214] (2) an injury to a person, property, or business; and (3) that the social media company knowingly provided substantial assistance, or conspired with the person who committed the terroristic acts.[215]

The ATA therefore contemplates two types of liability: direct liability and aiding and abetting liability.[216] As the courts have explained it, direct liability occurs because under 18 U.S.C. § 2339B(a)(1), it is a crime to provide material support to

[210] 18 USC § 2331.

[211] Jamie L. Boucher, Eytan J. Fisch, Ryan D. Junck, Margaret E. Krawiec & Timothy G. Nelson, *The Potential Impact of Terrorism Lawsuits Under the Antiterrorism Act on Ordinary Corporate, Banking and Sovereign Enterprises*, SKADDEN (May 26, 2020), https://www.skadden.com/insights/publications/2020/05/the-potential-impact-of-terrorism-lawsuits [https://perma.cc/7W2A-72Y8].

[212] *Id.*

[213] 18 USC § 2333.

[214] Under the ATA, international terrorism is defined as "activities that—(A) involve violent acts . . . that are a violation of the criminal laws of the United States or of any State, . . . (B) appear to be intended—(i) to intimidate or coerce a civilian population; (ii) to influence the policy of a government by intimidation or coercion; or (iii) to affect the conduct of a government by mass destruction, assassination, or kidnapping; and (C) occur primarily outside the territorial jurisdiction of the United States, or transcend national boundaries False." 18 USC § 2331. As such, acts of domestic terrorism (such as the January 6, 2021, attack on the U.S. Capitol) would not qualify even if those same acts—had they occurred abroad—would likely fit within the definition. *See* Del Quentin Wilber, *FBI Director Says Capitol Riot Was 'Domestic Terrorism,'* L.A. TIMES (March 2, 2021, 2:49 PM), https://www.latimes.com/politics/story/2021-03-02/fbi-wray-testify-congress-capitol-siege [https://perma.cc/FSH5-T3CT]; *see also Examining the Domestic Terrorist Threat in the Wake of Attack on the U.S. Capitol: Hearing before the H.R. Comm. on Homeland Sec.*, 117th Cong. (2021), https://www.congress.gov/event/117th-congress/house-event/LC65965/text?q=%7B%22search%22%3A%5B%22capitol+attack%22%5D%7D&s=1&r=7 [https://perma.cc/9G4U-7ZVW].

[215] 18 USC § 2333.

[216] In that sense, the way that the ATA establishes liability (i.e., under either a direct or secondary liability framework) is similar to how agency principles establish liability for the principal based on the agent's actions. *See* discussion *supra* Section III.A. As a result, plaintiffs would need to discuss primary and secondary liability at two different stages of the litigation: initially to establish whether the algorithm's actions were sufficient to establish a cause of action under the ATA (either direct or aiding and abetting liability) and then, subsequently, using common law theories of either direct or vicarious liability to hold *Facebook* liable for the algorithm's acts. *See* discussion *infra.*

a foreign terrorist organization. And if that material support also qualifies as an act of "international terrorism" under 18 U.S.C. § 2331(1), a plaintiff can recover for injuries that occurred "by reason of" the defendant's conduct.[217]

So, for instance, if a U.S. person were to fund and organize a U.S. cell of a foreign terrorist organization (as defined by the statute),[218] then, presumably, this would fit within the framework of the ATA.[219]

a. Direct Liability Under the ATA

Based on the current jurisprudence related to social media companies, a successful cause of action against Facebook based on direct liability for the ATA would remain unlikely, even after Section 230 reform. Specifically, direct liability under the ATA includes a causation requirement; plaintiffs must show that the defendants' actions proximately caused the harms alleged.[220] At least one court has held that "in order to determine proximate cause under the ATA, substantiality, directness, and foreseeability are all relevant."[221] As the Sixth Circuit put it when dismissing an ATA claim against Twitter, "this theory of direct liability requires Plaintiffs to show that Defendants, by providing social media platforms to ISIS, committed an act of international terrorism."[222] Therefore, under the current

[217] Crosby v. Twitter, Inc., 921 F.3d 617, 622 (6th Cir. 2019).

[218] As discussed *infra*, the foreign terrorist organization (FTO) requirement provides an additional hurdle for plaintiffs. Rather than simply proving that the act of terror was committed, plaintiffs must also show that the act of terror was committed by an FTO. Moreover, the FTO status comes as a result of a government designation rather than a post hoc analysis by the court. For a comprehensive discussion on the FTO requirement, see Dale Kim, Note, *The Inadequate Reach of Aiding and Abetting Liability under the Antiterrorism Act*, 59 COLUM. J. TRANSNAT'L L. 209 (2020).

[219] For instance, the statute was used to hold a U.S. bank liable for providing substantial assistance to a terrorist organization. *See* Linde v. Arab Bank, PLC, 882 F.3d 314, 318 (2d Cir. 2018).

[220] *Crosby*, 921 F.3d at 622–23.

[221] *Crosby*, 921 F.3d at 624. This test lines up with how the Supreme Court has articulated the standard. As the court noted in *Bridge v. Phoenix*, "[p]roximate cause . . . is a flexible concept that does not lend itself to 'a black-letter rule that will dictate the result in every case.' Instead, we 'use[d] 'proximate cause' to label generically the judicial tools used to limit a person's responsibility for the consequences of that person's own acts,' . . . with a particular emphasis on the 'demand for some direct relation between the injury asserted and the injurious conduct alleged.'" Bridge v. Phoenix Bond & Indem. Co., 553 U.S. 639, 654 (2008) (first quoting Holmes v. Sec. Inv. Prot. Corp., 503 U.S. 258, 272 n.20 (1992); then quoting *id.* at 268; and then citing Anza v. Ideal Steel Supply Corp., 547 U.S. 451, 461 (2006)).

[222] *Crosby*, 921 F.3d at 622. Interestingly, the Sixth Circuit explicitly declined to answer the question of whether "providing routine social media services could qualify as an act of international terrorism," *though* the court certainly seems skeptical of this possibility. *Id.* at n.2 ("We are making a big assumption here. For Defendants' conduct to qualify as an act of international terrorism, Plaintiffs must establish that providing routine social media services involve 'violent acts or acts dangerous to human life,' . . . are intended to 'intimidate or coerce' civilians, influence government policy through 'intimidation or coercion,' or affect the government through 'mass destruction, assassination, or kidnapping,' . . . and must 'transcend national boundaries.' *Id.* § 2331(1)(C). Any one of which would be a substantial hurdle for Plaintiffs.") Some scholars believe that the proximate cause element might prove to be an almost insurmountable hurdle. *See, e.g.*, Ellen Smith Yost, Note, *Social Support for Terrorists: Facebook's "Friend Suggestion" Algorithm, Section 230 Immunity, Material Support for*

case law, in order to establish a direct liability claim for a social media company like Facebook, plaintiffs would have to show that the company's algorithm committed an act of terrorism when it either (1) promoted content that directly led to terrorist activities or (2) provided friend recommendations to an account associated with a terrorist organization that substantially led to a foreseeable act of terrorism. As one scholar noted, "[t]hough no court has yet considered this issue, it will be very difficult for plaintiffs bringing algorithmic social media material support claims to meet this proximate cause standard."[223] Only a scenario with very specific connections between the terrorism and the algorithm's actions is likely to be successful.

b. Secondary Liability Under the ATA

Aiding and abetting liability under the statute is more attenuated and potentially even harder to prove. For instance, aiding and abetting liability first requires that a plaintiff show: "(1) the party whom the defendant aids . . . perform[ed] a wrongful act that causes an injury; (2) the defendant [was] generally aware of his role as part of an overall illegal or tortious activity at the time that he provides the assistance; [and] (3) the defendant . . . knowingly and substantially assist[ed] the principal violation."[224] In addition, as the Sixth Circuit in *Crosby* noted: "secondary liability requires that an act of international terrorism was 'committed, planned, or authorized' by a "foreign terrorist organization."[225] Once that has been established, plaintiffs can prevail if they show that the defendants' act aided and abetted the terrorist conduct.[226]

However, as the *Crosby* court implied, a mere allegation that a user was exposed to radicalized conduct due to algorithmic amplification is unlikely to be successful under the ATA.[227] Indeed, short of a showing that the algorithmic amplification affirmatively assisted the terrorist's actions, a cause of action under the ATA is unlikely to stand. [228] The *Crosby* court did, however, leave open the

Terrorists, and the First Amendment, 37 SANTA CLARA HIGH TECH. L.J. 301, 324–25 (2021) (discussing the hurdle posed by proximate cause elements for an algorithm to provide material support).

[223] Yost, *supra* note 222, at 324.

[224] *See* Halberstam v. Welch, 705 F.2d 472, 477 (D.C. Cir. 1983). It was this framework that Congress specifically endorsed when it passed amendments to the ATA allowing for aiding and abetting liability. *See* Kim *supra* note 218, at 224.

[225] *Crosby*, 921 F.3d at 626 (citing 18 U.S.C. § 2333(d)(2) (as designated under 8 U.S.C. § 1189)).

[226] In enacting amendments to the ATA that established secondary liability, Congress explicitly included corporations as possible defendants in this new framework. Kim, *supra* note 218, at 225. However, even with the amendments to the ATA, secondary liability against corporations has been difficult to prove. One early victory for plaintiffs was *Linde v. Arab Bank, PLC*, 882 F.3d 314, 318 (2d Cir. 2018), which led to a settlement agreement by the parties (although the district court's opinion was vacated on other grounds). *Id.* at 325. However, despite the rise in ATA claims, one commentator has noted that "[n]one of these [claims] has, to date, resulted in a final ATA judgment against a private company." Boucher et al., *supra* note 211.

[227] *Crosby*, 921 F.3d at 625.

[228] *Id.* at 624–25.

possibility of future ATA claims: "to be sure, this does not mean that Defendants could never proximately cause a terrorist attack."[229] Accordingly, a fact pattern like the hypothetical above may have a sufficient nexus between algorithmic act and liability to overcome the hurdle in *Crosby*. However, the evidence to establish causation would need to be substantial.

Algorithm as Agent for ATA

Specifically, if in our hypothetical plaintiffs establish that the genocidal conduct was organized by a "foreign terrorist organization" (FTO)—which is supported by the facts of the hypothetical here—then plaintiffs could use Facebook's algorithmic amplification as a predicate fact that would help to establish that the company aided and abetted the violence. In contrast to the allegations alleged in *Crosby*, these facts seem easier to establish as a claim: experts could be brought in to show that amplifying an FTO's violent rhetoric and actions in an unstable region like Qumar "substantially assisted" the FTO in their acts of terrorism. For instance, if the amplification directly led to an uptick in recruitment, that could be one way to meet the necessary aiding and abetting standard. A showing that the virality of the posts (also thanks to the algorithm) emboldened the FTO to engage in still more acts of violence might be another way to establish secondary liability.

Hypothetical Facebook Liability

If plaintiffs establish during discovery the facts discussed above, they could then use agency principles to establish liability against the company. For instance, as discussed above, if part of the reason the recommendation algorithm promoted the radicalized content is because of the company's negligent design or supervision (either by the developers of the algorithm or the integrity teams or data scientists employed by Facebook),[230] then liability could be established. Similarly, if plaintiffs can establish that Facebook knew that its algorithms were amplifying the terrorist conduct and nonetheless continued to allow the algorithm to be deployed, this could be used to show that Facebook ratified the algorithm's acts.[231] Finally, plaintiffs could likely introduce evidence showing that Facebook "controlled" every aspect of the algorithm, thus establishing sufficient control that would, in turn, lead to a showing of vicarious liability for Facebook even if the company was unaware of the impact its algorithm was having in Qumar.[232]

[229] *Id.* at 625.

[230] Among the evidence that plaintiffs would want to introduce are documents relating to whether and how the designers continued to monitor the algorithm's outputs once it was deployed.

[231] Another way that Facebook's liability could be established is by demonstrating that if the "principal" had engaged in the same conduct as its algorithm (in this case, amplifying posts) it would have been considered tortious conduct.

[232] For a discussion of the ATA within the context of social media companies under the current 230 regime, see Jaime M. Freilich, Note, *Section 230's Liability Shield in the Age of Online Terrorist Recruitment*, 83 BROOK. L. REV. 675 (2018).

3. Tort Claims

As we mentioned at the outset, there are a number of claims under common law tort causes of actions that plaintiffs in a post-Section 230 world could use to establish liability for social media companies. With the exception of claims for defamation, explored below, plaintiffs who allege intentional torts (such as assault, fraud, etc.) will face significant hurdles in successfully litigating their claims.[233] However, plaintiffs who use negligence as the basis for their tort claims may be more successful, and below, we discuss claims for negligent infliction of emotional distress. Nonetheless, even for claims that sound in negligence, there may be challenges to successfully litigating these lawsuits against a social media company, including a lack of precedent holding companies responsible for algorithmic acts. Establishing traditional tort law concepts like "duty"[234] and "causation" might also prove daunting.

To help illustrate what these claims might look like, and what challenges they might face, imagine a hypothetical inspired by recent headlines:[235] a post to Facebook announces that a major online furniture store is involved in a child sex trafficking ring, using astronomical prices for their furniture as coded messages to advertise the children who are "available." For instance, the post claims that a couch labeled as "the Tabitha," which is selling on the site for $10,000, is actually a hidden message that the company is using its site to engage in the trafficking of a girl with the same name. The post includes a picture of a girl, purportedly missing, who is also named Tabitha.

The post goes viral, in part because Facebook's recommendation algorithm promotes it to the top of users' feeds, and is viewed by millions of users. Despite the fact that the story is repeatedly and quickly debunked by several news outlets, it continues to circulate widely on Facebook. Several Facebook users recognize the girl in the picture (who is actually safely at home with her parents) and send both her and her parents Facebook messages, text messages, and posts on other social media outlets repeating the charges and inquiring into whether "Tabitha" is ok. "Tabitha" streams live on Facebook, insisting that she is ok and not being trafficked, but commenters refuse to believe her.

[233] While case law does provide for vicarious liability for intentional torts, *see, e.g.*, Manning v. Grimsley, 643 F.2d 20 (1st Cir. 1981), those causes of actions are outside the scope of this Article.

[234] Other scholars have already begun to analyze and assess other elements of a negligence claim. *See e.g.*, Chagal-Feferkorn, *supra* note 148 (examining the reasonable person standard needed for the breach element in a negligence claim specifically within the context of the "reasonable algorithm"); Vladeck, *supra* note 19, at 144 (discussing the standard of "reasonable decision making" for autonomous machines like self-driving cars); Weston Kowert, Note, *The Foreseeability of Human-Artificial Intelligence Interactions*, 96 TEX. L. REV. 181 (2017) (discussing the foreseeability/proximate cause element of a negligence claim, specifically within the context of humans interacting with AI systems); Selbst, *supra* note 19 (using the foreseeability dimension of negligence law to challenge the "conventional wisdom" that a negligence framework would work with AI).

[235] This hypothetical is based on reporting from *The Washington Post*, discussing a QAnon conspiracy theory about an alleged child sex trafficking ring. *See* Contrera, *supra* note 188.

In addition, as a result of the posts, the FBI, police, and organizations that combat trafficking are overwhelmed with calls asking them to investigate the furniture company. As a result of the publicity and social media attention, Tabitha refuses to leave her house and begins to suffer symptoms of trauma—worried that she might in fact be kidnapped at any moment. Her parents are also devastated. What might tort claims against a social media company like Facebook look like in such a hypothetical, were they not barred by Section 230?

a. Defamation Claims

Plaintiffs have always been able to sue individuals who defame them on social media sites, since Section 230 only shields the sites themselves from liability; the *speakers* of the allegedly defamatory content remain liable.[236] For instance, one of Elon Musk's posts on Twitter, a tweet he posted prior to buying the company in 2022, led to a civil lawsuit asserting defamation that Musk successfully defended and, in the words of one writer, also likely re-wrote defamation law in the process.[237] But Section 230 has been consistently held to block all defamation claims against social media companies hosting those same user's posts. As the Second Circuit has held, "[a]t its core, § 230 bars 'lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions—such as deciding whether to publish, withdraw, postpone or alter content.'"[238]

But what would liability look like in a post-Section 230 world? Would companies be held liable for defamation that they help promote through affirmative decisions to recommend the defamatory content to their users? If Section 230 is amended to allow for claims of defamation against the social media companies themselves, the basic elements of a defamation claim will remain unchanged. Defamation is "[a]n act of communication (whether written or oral) that tends to damage another's reputation to the extent of lowering their regard in the community or deterring others from associating with them."[239] Specifically, the Restatement provides four elements for a claim:

(a) a false and defamatory statement concerning another;
(b) an unprivileged publication to a third party;
(c) fault amounting at least to negligence on the part of the publisher; and

[236] *See, e.g.*, Jackson v. Mayweather, 10 Cal. App. 5th 1240 (2017) (ex-girlfriend of boxer Floyd Mayweather claimed that he had defamed her in connection with statements he posted on Facebook related to their relationship).

[237] Tom Hals, *Musk's Defamation Win May Reset Legal Landscape for Social Media*, REUTERS (Dec. 6, 2019, 6:03 PM), https://www.reuters.com/article/us-musk-lawsuit-landscape/musks-defamation-win-may-reset-legal-landscape-for-social-media-idUSKBN1YB023 [https://perma.cc/DC6S-FF4M] (noting that Musk's victory in the lawsuit may indicate that jurors believe that social media posts may be subject to a higher standard of liability). That claim was directed at Musk alone and not at Twitter for amplifying the allegedly defamatory content.

[238] Fed. Trade Comm'n v. LeadClick Media, LLC, 838 F.3d 158, 174 (2d Cir. 2016) (quoting Jones v. Dirty World Ent. Recordings LLC, 755 F.3d 398, 407 (6th Cir. 2014)).

[239] RESTATEMENT (SECOND) OF TORTS GLOSSARY DEFAMATION (AM. L. INST. 1977).

(d) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.[240]

Algorithm as Agent for Defamation

Regarding the hypothetical above, the false claims regarding "Tabitha" would certainly fit as "false and defamatory" statements that were published to a third party via Facebook. The plaintiffs would likely allege that Facebook's recommendation algorithms helped fuel the virality of the post by prominently positioning it in people's newsfeeds or potentially even recommending it to users. Allegations like these could well form the basis for "fault amounting at least to negligence on the part of the publisher."[241] It is unlikely that Facebook would be able to defend such a claim by saying that it was the algorithm, rather than Facebook agents, who made the decisions to promote the post. As mentioned in Section I.D, any recommendation algorithms used to promote posts like our hypothetical one are designed by humans, reflect human decisions, and are intentionally deployed to optimize some objective (here, user engagement with the site). Recent revelations from the Facebook Papers suggest that the company was arguably on notice that extremist content was the most likely to go viral,[242] and so continuing to use their recommendation algorithms could amount to at least negligence and meet the standard for defamation.

Hypothetical Facebook Liability for Defamation

Once algorithmic liability has been established, plaintiffs could proceed based on theories of both direct and vicarious liability to attempt to hold Facebook liable. As discussed above, a cause of action could stand if plaintiffs establish that Facebook knew of the virality of the false posts and did nothing to take it down (thereby ratifying the content). Similarly, a cause of action for vicarious liability could stand if Facebook was shown to have exerted significant control over the algorithm and the algorithm performed as designed.

Facebook could be especially vulnerable to defamation claims made for content that was shared by millions of users who were excluded from standard content moderation through its "CrossCheck" program, described above. "Under the program, those users are 'whitelisted'—rendered immune from enforcement actions—while others are allowed to post rule-violating material pending Facebook

[240] *Id.* § 558.

[241] While a full analysis of this issue is outside the scope of the article, it is worth noting that there is a distinction here regarding how defamation suits characterize "publisher" under common law torts and how Section 230 defines and uses the word "publisher." As such, litigants would want to take care to use the distinction carefully.

[242] *See e.g.,* Hagey and Horwitz, *supra* note 69 (describing leaked Facebook documents where company researchers noted that "[m]isinformation, toxicity, and violent content are inordinately prevalent among reshares," and also describing an email from a news publisher to a Facebook official indicating that "most divisive content that publishers produced was going viral on the platform").

employee reviews that often never come."[243] In other words, the posts of users in the CrossCheck program are left up longer, even when they violate Facebook's community standards. This decision and the fact that it might provide greater reach for defamatory content could have an impact on whether Facebook was negligent for purposes of a defamation claim. Further, although Facebook did not moderate the content of these users in the standard way, it appears that the site's recommendation algorithms did still amplify the posts (as evidenced by the virality of certain posts).

b. Negligent Infliction of Emotional Distress (NIED)

Scholars have long debated the contours of a claim for NIED,[244] and have disagreed about whether it is an independent tort or a subcategory of other negligence claims.[245] However, as the Supreme Court has stated, "[n]early all of the States have recognized a right to recover for negligent infliction of emotional distress, as we have defined it."[246] For those courts that do recognize the independent claim, they have established the elements as follows: "a legal duty of the defendant to protect the plaintiff from injury,"[247] a breach of that duty,

[243] Horwitz, *supra* note 37.

[244] *See, e.g.*, Virginia E. Nolan & Edmund Ursin, *Negligent Infliction of Emotional Distress: Coherence Emerging from Chaos*, 33 HASTINGS L.J. 583 (1982). In this article, the authors examine the California Supreme Court's conflicting decisions in *Dillon v. Legg*, 441 P.2d 912 (Cal. 1968), *Justus v. Atchison*, 565 P.2d 122 (Cal. 1977) (in bank), *disapproved of by* Ochoa v. Superior Court, 703 P.2d 1 (Cal. 1985), and *Molien v. Kaiser Foundation Hospitals*, 616 P.2d 813 (Cal. 1980) (in bank). *Dillon*, in particular, was a landmark case. The authors noted that the court listed three factors in assessing whether the defendant's emotional injury was foreseeable to the plaintiff: "(1) whether plaintiff was located near the scene of the accident . . . ; (2) whether the shock resulted from a direct emotional impact upon the plaintiff from [observing] the accident or from learning of the accident from others after its occurrence; and (3) whether the plaintiff and the victim were closely related . . . ." Nolan & Ursin, *supra*, at 588-89; *see also* Robert J. Rhee, *A Principled Solution for Negligent Infliction of Emotional Distress Claims*, 36 ARIZ. ST. L.J. 805 (2004). California has one of the most expansive jurisprudences on the subject and, in its leading cases, holds that physical injury is not necessary for a claim on Negligent Infliction of Emotional Distress ("NIED") to stand. *Molien*, 616 P.2d at 817. *But see* Thing v. La Chusa, 771 P.2d 814, 815 (Cal. 1989) (in bank) (holding that "in the absence of physical injury or impact to the plaintiff himself, damages for emotional distress should be recoverable only if the plaintiff: (1) is closely related to the injury victim, (2) is present at the scene of the injury-producing event at the time it occurs and is then aware that it is causing injury to the victim and, (3) as a result suffers emotional distress beyond that which would be anticipated in a disinterested witness.").

[245] Gregory C. Keating, *Is Negligent Infliction of Emotional Distress a Freestanding Tort?*, 44 WAKE FOREST L. REV. 1131 (2009). Keating argues that yes, NIED is a freestanding tort. *Id.* at 1136. Specifically, Keating argues that NIED liability should be classified as a doctrine of proximate cause rather than a doctrine of duty. *Id.* The purpose of NIED as a tort is to "protect[] people's interest in 'emotional tranquility'" and because "NIED shatters a plaintiff's tranquility through conduct that is inadvertent, inattentive, or otherwise insufficiently careful," proximate cause is the appropriate analysis. *Id.* at 1138–39.

[246] Consol. Rail Corp. v. Gottshall, 512 U.S. 532, 545 (1994) (providing extensive discussion of NIED claims based on emotional distress and concluding that a plaintiff must show that they were in the "zone of danger" created by the defendant's negligent conduct to recover).

[247] Couzens v. Donohue, 854 F.3d 508, 518 (8th Cir. 2017) (quoting Thornburg v. Fed. Express Corp., 62 S.W.3d 421, 427 (Mo. Ct. App. 2001)).

proximate cause, and injury—as well as two additional elements—"that the defendant should have realized that his conduct involved an unreasonable risk of causing distress"[248] and "that the emotional distress or mental injury must be medically diagnosable and must be of sufficient severity so as to be medically significant."[249]

Because courts have consistently dismissed claims brought against social media companies for NIED at the motion to dismiss stage, no court has had to grapple with how the tort might map on to hypotheticals like ours. It is difficult to know, for example, whether a plaintiff would be able to establish that a social media company like Facebook owed them a duty, as the tort requires. If Section 230 is amended and the claims advance past the motion to dismiss stage, plaintiffs may use theories regarding special relationships between corporations and the individuals that they interact with to establish potential liability under these claims, even in instances where corporations allege nonfeasance.[250]

Algorithm as Agent for NIED

Assuming plaintiffs can show a duty by Facebook to its users, a cause of action for NIED may be successful if the company's recommendation algorithm can be shown to directly cause the plaintiff's emotional distress. In our hypothetical, "Tabitha" could argue that it was not the original post but rather *the fact that it went viral*—something caused at least in part by Facebook's recommendation algorithm—that caused her harm.

Further, Tabitha's parents might have an NIED claim as well. The Restatement provides that close family members of victims who experience serious bodily injury may bring a claim for the emotional harm the close family members experienced watching their loved one go through the ordeal.[251]

[248] *Id.*

[249] *Id.*

[250] Right now, however, using a special relationship to establish a corporation's duty is largely theoretical. Nonetheless, within the field of business and human rights, scholars have begun to examine the issue. *Cf.* Jena Martin Amerson, *What's in a Name? Transnational Corporations as Bystanders under International Law*, 85 ST. JOHN'S L. REV. 1 (2011) (analyzing potential legal frameworks for holding corporations accountable as bystanders); Jena Martin Amerson, *"The End of the Beginning?": A Comprehensive Look at the U.N.'s Business and Human Rights Agenda from a Bystander Perspective*, 17 FORDHAM J. CORP. & FIN. L. 871, 885 (2012) (discussing nonfeasance tort liability and the special relationship needed for corporate bystanders); Gwynne Skinner, *Rethinking Limited Liability of Parent Corporations for Foreign Subsidiaries' Violations of International Human Rights Law*, 72 WASH. & LEE L. REV. 1769, 1786 (2015).

[251] 2 RESTATEMENT (THIRD) OF TORTS—LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 48 (AM. L. INST. 2012). According to § 48, an actor who negligently causes sudden serious bodily injury to a third person is subject to liability for serious emotional harm caused thereby to a person who (a) perceives the event contemporaneously, and (b) is a close family member of the person suffering the bodily injury. *Id.*

Hypothetical Facebook Liability

If our hypothetical were extended to a frightening but not implausible scenario—where a person viewed the Facebook post about Tabitha and tried to "rescue" her by taking her from her parents—serious bodily injury to her could result.[252] As noted above, an element of an NIED claim is that the defendant must have realized that its conduct caused an unreasonable risk of causing distress. Again, the revelations of the Facebook Papers would prove potentially helpful to plaintiffs on this point, since they suggest that the company was on notice that its recommendation algorithm was pointing people toward more extreme and divisive content.[253]

Either of these cause of actions (as well as others)[254] may be brought under common law torts.

IV. Considerations For Congress and Courts

In the previous Sections of this article, we have tried to present specific, tangible frameworks that judges and practitioners could use in the (likely) event that Section 230 protections are limited (either through legislative or judicial action). However, we also recognize that, as these are all matters of first impression, both courts and legislators will have to wrestle with the underlying policy considerations that affect their decisions in such a dominant area of commercial liability. We do not have all the answers. However, in this Section, we attempt to provide a very brief overview of some considerations that lawmakers and courts will need to engage with as they map this new and developing legal terrain. These questions are

[252] It is also possible that a court could find that Tabitha's trauma amounted to a serious bodily injury produced by the emotional distress she experienced.

[253] *See supra* Part I, examining the *Wall Street Journal's* reporting on Facebook data scientists and the company's "increasing liability" for its recommendation algorithm.

[254] As the analysis above makes clear, there are significant legal hurdles for tort claims against social media companies for claims of defamation and NIED, even once Section 230 is amended. Those hurdles would be even larger for a claim of wrongful death, and so we only briefly address it here. There are a number of underlying facts that can lead to a wrongful death claim. As such, although the cause of action itself would be labeled "wrongful death" the underlying facts can arise out of either negligent or intentional conduct. For instance, cyber-bullying can escalate to physical confrontation that can then lead to murder. In that case, and in addition to whatever charges the government may bring under criminal law, the victim's estate can bring a civil cause of action against the perpetrator for wrongful death. More often, however, wrongful death claims arise when the underlying facts sound in negligence rather than in intentional tort. Given what we know about how Facebook's algorithm operates—amplifying divisive and emotionally charged content—then it would seem logical to assume that the next step in the lawsuit progression is a cause of action that claims that this amplification caused a death—either through suicide or murder. Any wrongful death claims would have a significant causality connection to overcome. However, once overcome, a wrongful death claim could stand. Although still a small part of the academic literature, there is a growing body of scholarship that does make the connection between social media and murders. *See, e.g.*, Brandy Nichole Jones, The Influence of Social Media on Murder (June 2020) (M.A. thesis, California State University, San Bernardino), https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2239&context=etd [https://perma.cc/BRH7-SJ7E].

primarily practical: What should reform of Section 230 look like? How should we craft a law (either in statute or judicially) that is concrete enough to assist practitioners now and yet flexible enough to adapt to the next technological innovation in this area? We provide our initial thoughts here, and note that each of these is an area that will benefit from further scholarly development.

A. For Lawmakers

As lawmakers craft bills to address algorithmic amplification and to alter in some way Section 230, they will be balancing many competing demands and attempting to avoid unintended consequences.[255] We offer here two brief recommendations. First, Section 230 should be amended to exclude from the liability shield affirmative decisions made to promote content, a la Judge Katzman's dissent in *Force*. There may be compelling reasons to allow social media companies to continue to benefit from a liability shield for problematic content a user posts and which the site promptly removes. But there should be no shield from a site's affirmative decision to promote content, whether by positioning it prominently in a user's newsfeed or by recommending that other users join a group based on that problematic content.

Second, lawmakers should look to existing and emerging best practices for content moderation and algorithmic amplification. For instance, lawmakers should consult existing scholarship that focuses on developing policies based on the likelihood of the underlying risk of harm[256] posed by an algorithm's objective. The field of ethical algorithmic design already has a robust body of scholarship that highlights many best practices.[257] Further, if they decide to regulate content moderation practices, lawmakers should look to existing industry standards, such as the Santa Clara Principles on Transparency and Accountability in Content Moderation.[258] Any guidelines developed should focus specifically on the importance of testing these algorithms prior to deployment, before they are unleashed on an unsuspecting public.[259]

[255] *See, e.g.*, Matt Perault, *Well-Intentioned Section 230 Reform Could Entrench the Power of Big Tech*, SLATE (June 1, 2021, 9:00 AM), https://slate.com/technology/2021/06/section-230-reform-antitrust-big-tech-consolidation.html [https://perma.cc/J8U8-DBZG].

[256] Heiss, *supra* note 20, at 195.

[257] *See, e.g.*, MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN (2019).

[258] *See, e.g.*, Jen Patja Howell, *The Lawfare Podcast: Working Toward Transparency and Accountability in Content Moderation*, LAWFARE (Dec. 23, 2021, 5:01 AM), https://www.lawfareblog.com/lawfare-podcast-working-toward-transparency-and-accountability-content-moderation [https://perma.cc/VZ6A-KGNW].

[259] One example can be found at New York University's AI Now Institute, which has "introduced a model framework for governmental entities to use to create algorithmic impact assessments (AIAs), which evaluate the potential detrimental effects of an algorithm in the same manner as environmental, privacy, data, or human rights impact statements." Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, Brookings Institute Report (May 22, 2019), https://www.brookings.edu/research/

Of course, even these modest suggestions for lawmakers lead to additional questions they will need to answer. For instance, if social media companies do, in fact, follow the regulatory best practices, should that provide a defense against liability? A rebuttable presumption that they were not negligent? What might a due diligence framework look like?[260] Should the companies' algorithms be audited by an agency, perhaps the FTC, in much the same way that FINRA audits sensitive financial documents to look for violations?[261] These intriguing and essential questions will require careful consideration from lawmakers, scholars, and jurists.

B. *And for Courts*[262]

If claims against social media companies are allowed to move forward in the absence of a Section 230 shield, we anticipate that there will be many, many such claims. We argue that claims premised on a theory that plaintiffs were harmed by algorithmic amplification deserve to reach discovery, and should not be automatically rejected at the motion to dismiss stage. But we also urge courts to take seriously the causation elements that are present in each of the potential claims we consider above. We believe that these causation elements—rather than Section 230—could serve as the gatekeeper to litigation. We are not alone in noting that claims premised on algorithmic acts will face complicated and significant causation issues.[263] We recognize that defending such claims through discovery will be expensive for defendants, but we believe this balance is the more appropriate one in light of serious concerns about the devastating impacts of algorithmic amplification.

algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ [https://perma.cc/E2EX-7MLB].

[260] To that end, lawmakers could look at the business and human rights framework for guidance. Specifically, the UN Guiding Principles on Business and Human Rights exhorts companies to engage in human rights due diligence, wherein companies assess the risk of harm for its action vis-a-vis the larger community (rather than simply assessing risk to the company's reputation). This approach has been taken up by countries in Europe. Rachel Chambers & Jena Martin, *Foreign Corrupt Practices Act for Human Rights: A Due Diligence Plus Model for the United States*? (W. Va. Coll. of Law, Research Paper Series No. 2021-019, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3852975 [https://perma.cc/BVE9-ZK7V]. Heiss also advocates for analyzing harm within the context of high-risk AI impacts and lower risks AI impacts. Heiss, *supra* note 20, at 195.

[261] Michael Kearns & Aaron Roth, *Ethical Algorithm Design Should Guide Technology Regulation*, BROOKINGS (Jan. 13, 2020), https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/ [https://perma.cc/4KY5-MXTQ] (advocating that the FTC audit algorithms in a manner similar to the FTC/FINRA).

[262] We undertake our preliminary analysis bearing in mind Matthew Scherer's admonition that the characteristics of a common law court system "make the tort system a mixed blessing when it comes to the management of public risks caused by emerging technologies." Scherer, *supra* note 19, at 389. Nevertheless, given the rise of cases (discussed above) alleging causes of actions based on AI, we believe that it would still be wise to provide some guidance to courts as they undertake the inevitable.

[263] *See, e.g.*, Heiss, *supra* note 20 ("New emerging technologies can raise sophisticated causation issues.").

CONCLUSION

Social media companies' use of algorithms to moderate, curate, and amplify user content presents a number of significant hurdles for practitioners, courts, and lawmakers to consider. We hope, however, that by providing an initial framework for examining agency relationships between the companies and the algorithms they deploy, we have provided one way forward for framing the issue. Ultimately, it will be a delicate balance to strike, as policy makers attempt to reign in the negatives associated with algorithmic amplification while not stifling innovation or removing the positive benefits associated with social media usage. A change is indeed coming, and it is imperative that we consider the full ramifications of any proposals. As Haugen noted, "a lot of lives [are] on the line."[264]

[264] *See* Scott, *supra* note 50.