

Enhancing the FTC's Consumer Protection Authority to Regulate Social Media Companies

BY PAUL M. BARRETT AND LILY WARNKE



NYU STERN

Center for Business
and Human Rights

February 2022

Table of Contents

Executive Summary	2
1. Introduction	3
2. General Principles	4
3. Policy Proposals.....	5
4. Conclusion	12

Authors

Paul M. Barrett is the deputy director of the NYU Stern Center for Business and Human Rights and an adjunct professor at the NYU School of Law.

Lily Warnke, a former fellow at the Center, contributed important research and analysis to this paper.

Acknowledgments

We are grateful to Craig Newmark Philanthropies and the Open Society Foundations for their continued support for our work on the technology sector.

NYU Stern Center for Business and Human Rights
Leonard N. Stern School of Business
44 West 4th Street, Suite 800
New York, NY 10012
+1 212-998-0261
bhr@stern.nyu.edu
bhr.stern.nyu.edu

© 2022 NYU Stern Center for Business and Human Rights
All rights reserved. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of the license, visit:
<http://creativecommons.org/licenses/by-nc/4.0/>.



To reduce the level of hate speech, health misinformation, incitement to political violence, and other forms of harmful content spread by social media platforms, Congress should pass legislation to achieve the following goals:

-
- Enhance the consumer protection authority of the Federal Trade Commission and direct the agency to conduct sustained oversight of social media companies.

 - Empower the FTC to enforce a new mandate that the companies maintain procedurally adequate content moderation systems, which deliver on promises made to users about platform rules and enforcement practices.

 - Direct the agency to enforce new transparency requirements, including that social media companies disclose how their algorithms rank, recommend, and remove content, as well as information about the kind of content that reaches large audiences and the factors that cause such content to “go viral.”

 - Oblige social media companies to maintain comprehensive, searchable advertising libraries, which disclose who has paid for each ad.

 - Respect the First Amendment by prohibiting the FTC from any involvement in substantive content decisions, including decisions to take down or leave up content.

 - Amend Section 230 of the Communications Decency Act of 1996 to clarify that its terms do not shield social media platforms from FTC enforcement actions or from private civil claims related to categories of particularly harmful content for which Congress determines platforms must be held accountable.

1. Introduction

Congress is considering dozens of bills designed to rein in social media platforms such as Facebook, Instagram, YouTube, and Twitter. The profusion of proposed legislation reflects a broadly held view that the companies that own these platforms have failed to self-regulate adequately. This failure has resulted in the amplification of misinformation, hate speech, incitement of political violence, and other harmful content. By intensifying the extreme polarization that now characterizes U.S. society, the platforms are undermining trust in elections and public health policies and blurring the distinction between provable facts and ideologically motivated falsehoods.

Remarkably, no arm of the federal government exercises sustained oversight of the social media industry in the way that the Federal Communications Commission regulates television and radio and the Securities and Exchange Commission oversees equity markets. In congressional testimony in the fall of 2021, Facebook whistleblower [Frances Haugen compared](#) the current regulatory environment for social media to “the Department for Transportation regulating cars by watching them drive down the highway.”

While most lawmakers apparently agree that something must be done about the malign effects of social media, there isn’t a consensus on how to proceed. The surfeit of legislative proposals reflects clashing motivations and competing regulatory strategies.¹

But even in what promises to be a rancorous midterm election year, there is an opportunity to move past overheated rhetoric and make progress on real reform. This paper does not endorse any specific legislation, although we refer to a number of bills that contain worthy ideas. The [Digital Services Oversight and Safety Act](#), for example, would create an appropriately staffed and funded digital bureau within the Federal Trade Commission to regulate the social media industry.

What follows are principles and policy goals intended to clarify the debate in Congress and shape an agenda for the FTC. We recommend that Congress direct the FTC to oversee the social media industry under the consumer protection authority that the agency already exercises in its regulation of other industries. Specifically, the agency ought to have the ability to require social media companies to maintain procedurally adequate content moderation policies and procedures, which deliver on promises made to users. Legislation also should authorize the FTC to require these companies to disclose information about how they rank, recommend, and remove content, as well as about the kind of content that reaches large audiences and the factors that cause such content to “go viral.” If applied prudently, this model for fostering transparency in the service of greater accountability could reduce the harms currently caused by a powerful and pervasive industry. All of this must be accomplished with due respect for the First Amendment’s prohibition of government speech restrictions.

Definitions

- **“Social media platforms”:** We are referring to social-networking sites like Facebook, Twitter, and Reddit, where users can post content and interact with other users; video-sharing services like YouTube and TikTok; and WhatsApp and other messaging services. We are *not* addressing search engines, such as Google, or online marketplaces, such as Amazon which, in our view, present distinctive problems that deserve separate legislative attention.
- **“Large platforms”:** Complying with some of our regulatory proposals would be onerous for smaller platforms with limited resources. To avoid inadvertently squelching innovation and further entrenching industry giants, Congress should limit application of these new regulations to large platforms, which lawmakers should define using a combination of factors, including number of monthly users, annual revenue, and workforce size.

¹ While grandstanding has eclipsed legislating in the U.S., the European Parliament in January approved a broad [Digital Services Act](#) that seeks to accomplish the kind of transparency and accountability we advocate in this paper. The DSA could take effect later in 2022, after negotiations between the Parliament, European Union member states, and the European Commission.

2. General Principles

Our policy proposals rest on the following principles:

- **Protecting free speech:** The First Amendment to the U.S. Constitution prohibits government restriction of social media users' online speech. It also forbids official interference with the choices social media companies make about which content to host—a protection similar to that enjoyed by a newspaper in its editorial choices. Our proposals are grounded in deep respect for this wise constitutional limit on government power. At the same time, the likelihood that even the most painstaking effort to regulate social media will elicit claims of First Amendment violations should not shut down discussion of potential reforms.
- **Imposing procedural, not substantive regulation:** To mitigate free-speech concerns, enhanced regulation should impose procedural, not substantive, obligations on social media companies. Apart from enforcing prohibitions on conspiracy, fraud, and other unlawful speech, the government should not make decisions about what content platforms host or amplify. Instead, legislation should require that if large platforms maintain content moderation systems, they include clearly explained rules, enforcement mechanisms, and means for appealing decisions. The platforms would continue to determine the substance of their content moderation systems. The FTC would ensure that once such systems exist, they have the human and automated resources needed to uphold the rules that platforms have promised users they will apply. In addition, the agency would oblige platforms to disclose key aspects not only of when and how they remove certain content, but also how they rank and recommend it for users.
- **Protecting consumer interests:** Building on the FTC's existing consumer protection authority offers the most promising strategy for reducing harms associated with social media.² [Section 5 of the Federal Trade Commission Act](#) provides the FTC with authority to stop acts or practices that are "unfair or deceptive" to consumers. Congress should pass legislation that defines the failure to maintain a procedurally adequate content moderation system as an unfair practice and therefore one that is unlawful. The legislation likewise ought to make clear that failure to disclose key aspects of the ranking, recommending, and removal of content constitutes a deceptive practice under Section 5. While U.S. legal concepts provide ample foundation for what we have in mind, the proposed [Digital Services Act](#) approved by the European Parliament in January 2022 provides complementary ideas about transparency and accountability.
- **Incentivizing industry vigilance:** It would have been preferable for social media companies themselves to have adequately policed the spread of harmful content, without government intervention. Voluntary self-regulation minimizes potential First Amendment complications. But the industry has not regulated itself sufficiently, necessitating action by Congress and the FTC. Still, the procedurally oriented consumer protection model we advocate will succeed only if social media companies commit to establishing and enforcing more stringent standards and procedures for removing and down-ranking harmful content. As a practical matter, only the companies themselves have the technological capacity to filter the billions of items their users post every day. Large mainstream tech companies, including Meta, which owns Facebook, Instagram, and WhatsApp, and Google, the owner of YouTube, already feature extensive moderation systems. Mainstream platforms know that without these systems they would be inundated by spam, pornography, hate speech, and other forms of unwelcome content that would drive away users and advertisers alike. To protect their business interests, the platforms will continue to moderate content on a large scale. The purpose of government oversight is to ensure that they do so more transparently and effectively.

² Mark MacCarthy, a senior fellow at the Institute for Technology Law and Policy at Georgetown Law School, has [trenchantly analyzed](#) the consumer protection approach to regulating content moderation.

Disclaimers

- **Limits of this paper:** Our proposal for enhanced FTC consumer protection does not address other potentially important avenues of reform. These include efforts to [update and strengthen antitrust enforcement](#), [shield user privacy](#), and [prevent various harms to children and teenagers](#). These topics deserve urgent attention and are the subject of a number of pending bills. But they are beyond the scope of this paper.
- **Fear of government overreach:** Skeptics of enhanced FTC oversight may raise the potential danger that a president could use the agency to punish social media companies and censor free speech. Would one want a politically appointed FTC chairman harassing Facebook, YouTube, or Twitter at the behest of a president who believes he has been wronged by the platforms? Not any more than one would want a vengeful president setting enforcement policy for the Internal Revenue Service, commanding the military, or exercising any of the other awesome powers entrusted to the nation's chief executive. But the possibility that regulatory authority, or other powers, might someday be abused can't be a determinative argument against providing a federal agency with the ability to accomplish its mission—in this case, consumer protection in a fast-evolving information marketplace.

3. Policy Proposals

The nature of the problem

All large tech platforms already publish extensive content moderation standards and policies for enforcing them. Unfortunately, the standards are both prolix and cryptic, often confusing, and sometimes down-right misleading. Enforcement ranges from overly broad to feeble to haphazard, especially in languages other than English.³

The sheer scale of social media traffic—billions of posts each day, in scores of languages—means that no system will provide perfect filtering. Some harmful content inevitably will surface and spread. But this problem of scale is one of the industry's making. In a short span of years, Mark Zuckerberg, the founder and CEO of Meta, and other social media moguls stressed growth at all costs without incorporating adequate safety precautions.

The consequences are now familiar: Facebook groups kindle baseless claims of a “stolen” election and calls for a violent revolt, fueling the [January 6 insurrection](#). YouTube provides “[one of the major conduits of online disinformation and misinformation worldwide](#),” amplifying hate speech against vulnerable groups, undermining vaccination campaigns, and propping up authoritarians from Brazil to the Philippines. Twitter fosters a “[disinformation-for-hire industry](#)” in Kenya, stokes civil war in Ethiopia, and spreads “fake news” in Nigeria.

We see the industry's malign effects, but we know very little about how its technology works or the choices company executives make. Social media companies offer transparency reports, but these turn out to be highly curated collections of [often-manipulated and unrevealing data points](#). Social media remains a black box.

Consider “engagement.” The major platform companies have consciously designed the algorithms that rank and recommend content to prioritize posts that users will “like,” share, and comment on. These posts are more likely to go viral and, in a self-reinforcing feedback loop, are prone to be recirculated and otherwise engaged with by larger audiences. As various studies have shown, [users tend to engage with](#)

³ The detailed [decisions of the Facebook Oversight Board](#) provide numerous illustrations.

content that elicits anger and fear. As a result, [ranking algorithms amplify this kind of sensationalistic and extreme content](#). Maximizing engagement increases partisan hatred, one group of researchers found, “[in part because of the contagious power of content that elicits sectarian fear or indignation](#).” Social media companies boost user engagement, not because they want to [intensify political polarization](#), but because the amount of time users spend on a platform interacting is also the amount of time they spend looking at the paid advertising that comprises the lion’s share of revenue for the major platforms.

“

Documents brought to light in the fall of 2021 by Facebook whistleblower Frances Haugen underscored that the platforms’ secrecy cloaks evidence that they are less effective than they would like to admit in filtering out harmful content.

”

But without access to the data, we don’t know the degree to which the algorithm designers and their corporate bosses seek to boost engagement. On rare occasions—for instance, during the attempt by former President Trump and some of his supporters to undermine the November 2020 election—tech companies have announced adjustments they’ve made to lower levels of fear and indignation among users. But when the immediate crisis recedes, the algorithmic dials typically are turned back to their normal positions to once again rev up engagement.⁴ Even when social media companies publicly acknowledge their ability to fine-tune the content mix, they closely guard the specifics, with corporate spokespeople citing the need to protect trade secrets from prying rivals.

Documents brought to light in the fall of 2021 by Facebook whistleblower Frances Haugen underscored that this secrecy cloaks evidence that the platforms are less effective than they would like to admit in filtering out harmful content. All of the major platforms ban dehumanizing attacks on people based on race, gender, religion, and certain other protected categories. Trying to demonstrate its vigilance, [Facebook boasts publicly](#) that more than 95% of the hateful content it takes down is removed by automated filtering before a user or paid human moderator notices it. But that impressive-sounding take-down rate means less than executives would have you believe. It applies not to *all* hate speech on the platform, but only to the offending content *that Facebook detects*. So, that’s 95% of—well, we have no idea, because Facebook doesn’t disclose the extent of hate speech on the platform. According to [one of Haugen’s leaked documents](#), the company’s own analysts “estimate that we may action as little as 3-5% of hate...on Facebook.” Little wonder that Facebook has stoked extremism and violence in countries ranging from the U.S. to [Myanmar](#) and [Ethiopia](#).

A realistic response to the problem

The most effective way to diminish social media’s tendency to amplify harmful content would be for the companies to replace their engagement-based business model. They could, for example, try selling subscriptions to users. But most users, conditioned to think of social media as “free,” probably would refuse to pay.⁵ Advertising would plummet; social media sites would shrivel.

Tech companies have shown no interest in sacrificing a large percentage of their revenue in a public-spirited attempt to reduce harmful content. And the government cannot force them to do so. The First Amendment would prohibit an order from Washington that social media companies change their fundamental strategy to alter the mix of content they host. Apart from free-speech concerns, a bipartisan fear of losing Silicon Valley financial support would make such an order politically infeasible.

Instead, what government can do is take measured steps under the familiar banner of consumer protection to shield social media users, and society at large, from the consequences of harmful content.

⁴ After the 2020 election, when President Trump and his supporters spread falsehoods about rogue ballots and fixed voting machines, Facebook made adjustments to calm the tone and content of users’ News Feeds. But in mid-November, a company vice president told journalists that the algorithmic changes would be reversed. See *An Ugly Truth* by New York Times reporters Sheera Frenkel and Cecilia Kang (<https://www.harpercollins.com/products/an-ugly-truth-sheera-frenkelcecilia-kang?variant=32999376551970>).

⁵ In fact, users already pay—with their attention, as well as personal data on buying tendencies, political leanings, and web-surfing history.

These real-life consequences include the vicious and sometimes violent conduct incited by hate speech; the erosion of trust in elections caused by political disinformation; and the mass illness and death, as well as economic and educational disruption, exacerbated by misinformation about pandemic-era public health policies.⁶

These steps would not eliminate the harmful side effects of social media use. The enormous scale of the major platforms makes it difficult to imagine, let alone implement, systems that would comprehensively prevent the spread of harmful content. What Congress can do is seek modest but meaningful reform by augmenting the FTC's consumer protection authority.

Transparency and accountability

It is difficult to chart a regulatory course in the dark. This remains true even in the wake of the Frances Haugen revelations. To paraphrase Donald Rumsfeld, when it comes to social media, there are things we know we don't know, and then there are the *unknown unknowns*—things we're not even aware we don't know. By requiring the industry to divulge more of its currently secret data—with stringent safeguards to protect user privacy and forestall government misuse of the information for surveillance purposes—regulation could help outside researchers better understand and describe the harmful effects of social media, which in turn, would assist policymakers in crafting more effective responses.

Voluntary attempts to facilitate independent social media research have sputtered. In one program known as [Social Science One](#), Meta gave researchers access to a large batch of internal information, only to announce in the summer of 2021 that the dataset had significant errors. In another incident last year, Meta cut off the personal accounts of NYU researchers and [shut down their study](#) of how misinformation spreads via political advertising on the platform.

Karen Kornbluh of the German Marshall Foundation and Ellen Goodman of Rutgers Law School have argued that, as a society, we need to [demand of social media what has been required of other industries](#) ranging from pharmaceuticals to airlines to chemicals: "These include responsibilities to disclose information, to demonstrate the safety of their products, to mitigate harm, and to comply with their own promises." The European Union, which is several steps ahead of the U.S. in formulating regulatory strategy for social media, has emphasized similar ideas in its proposed [Digital Services Act](#). The DSA would oblige large platforms to reveal more about their content-recommendation systems, allow independent audits of their risk-management efforts, and share key data with outside researchers, among many other provisions.

Fortifying the FTC's consumer protection authority

The FTC already has experience investigating unlawful practices by major social media companies and enforcing settlements requiring changed corporate behavior and payment of penalties. These cases generally have grown out of one-off investigations of privacy violations, rather than as part of a systematic exercise of industry-wide oversight authority.

In 2019, Facebook, now part of Meta, agreed to pay \$5 billion and change certain company practices to [settle FTC charges](#) that it violated an earlier agency order and deceived users about their ability to control the privacy of their personal information. The case grew out of the scandal involving Cambridge Analytica, a British data analytics firm that Facebook admitted had improperly harvested and saved user information. In a separate 2019 settlement, Google agreed to pay \$170 million to [resolve allegations by the FTC](#) and the New York Attorney General that its YouTube subsidiary illegally collected personal information from child users.

⁶ The term "disinformation" refers to intentionally false or misleading information intended to sow divisiveness and confusion. "Misinformation" refers to such content created or spread by people who may not understand that it is untrue. In either case, the false content may be intermingled with truthful statements, making it seem more credible.

Expanding its reach beyond these privacy cases, the FTC should rely on its consumer protection authority to oversee the social media industry on an ongoing basis. To clarify the appropriateness of this application of Section 5 of the FTC Act, and to strengthen the agency's position in federal court litigation, Congress needs to direct the agency explicitly to use its consumer protection authority in this manner.

The FTC already has experience investigating unlawful practices by major social media companies and enforcing settlements requiring changed corporate behavior and payment of penalties.

Section 5(a) protects consumers by providing that “unfair or deceptive acts or practices in or affecting commerce” are unlawful. An unfair act is one that “causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition.” Deceptiveness involves “a material representation, omission, or practice that is likely to mislead a consumer acting reasonably in the circumstances.” In a consumer protection ruling in 2015 that favored the FTC, the [U.S. Court of Appeals for the Third Circuit](#) observed that, as a practical matter, “facts relevant to unfairness and deception frequently overlap.”

To further fortify the FTC’s authority, Congress should clarify how the agency’s enhanced authority comports with existing elements of a

Section 5(a) violation. The “substantial injury” in question is users’ unwelcome exposure to harmful content that has real-world consequences: violence based on religious animus, racial discrimination, doubt cast on elections, misinformation that erodes public health policies, and so forth. This kind of substantial injury is not “reasonably avoidable,” short of consumers avoiding social media entirely.⁷

Addressing the possibility of “countervailing benefits to consumers or to competition,” Congress could concede that some consumers might prefer digital communication unfettered by content moderation, but the sort of restrained FTC oversight proposed here would not impinge on this interest. The government would not propose substantive content rules or enforce such rules; these would remain under the control of the social media companies. The government’s purview would be strictly limited to supervising procedural adequacy and data disclosure. As for the effects on competition, the safe harbor provisions discussed earlier in this paper, which would limit the reach of regulation to companies whose revenue, user headcount, and workforce exceed certain thresholds. This would protect against inadvertent economic harm to smaller rivals of the large platforms.

In similar fashion, Congress ought to underscore that, while the FTC has traditionally focused on marketplace “injuries” measured in dollars lost or physical harm, the analysis for social media needs to be different. It should encompass more intangible forms of injury—such as, psychological trauma and erosion of free and fair elections—which are associated with an unusually pervasive form of technology.

Although platform companies don’t receive monetary payment from users, there is a merchant-consumer relationship justifying application of consumer protection law. “Facebook takes its users’ time, attention, and personal data; in exchange, Facebook gives those users access to and use of Facebook’s social network, which Facebook itself deems a ‘product.’” This apt analysis came from a [trio of consumer-advocacy organizations](#) in a December 2021 friend-of-the-court brief supporting a pending lawsuit against Facebook in the Superior Court of the District of Columbia.⁸

If Congress determines that a broader or different definition of unfairness or deceptiveness under Section 5(a) is needed to undergird specific statutory authority for FTC regulation, lawmakers could expand the reach of these concepts based on the novel challenges posed by social media.

⁷ It’s worth pointing again to Mark MacCarthy’s [paper on regulating content moderation](#).

⁸ The pro-consumer groups Consumer Reports, Public Knowledge, and Upturn filed their amicus brief in support of Muslim Advocates, an organization that sued Meta.

Applying consumer protection standards to social media platforms

To oversee new requirements concerning social media, Congress should create a Bureau of Digital Services Oversight and Safety within the FTC and appropriate annual funding sufficient for the bureau to undertake its challenging mission.⁹ The FTC is relatively small and already stretched to the limit by the demands of its existing antitrust and consumer protection duties, including its ongoing [monopolization lawsuit against Facebook](#).¹⁰ To respond to the inevitable industry legal challenges to new regulation and to ensure the effectuation of congressional intent, the FTC's budget must expand. President Biden's stalled Build Back Better bill contained [\\$500 million in additional annual FTC spending](#). Something in that ballpark seems appropriate, given the scale of the industry and its negative effects on consumers.

The FTC's digital bureau would oversee the procedural adequacy of content moderation and new data-disclosure requirements:

Procedurally adequate content moderation: Congress should define procedurally adequate content moderation as a system based on clearly articulated rules, enforcement practices, and mechanisms for user appeals. Failure to meet these requirements, as determined by the FTC, would constitute unfair or deceptive acts or practices and be grounds for agency enforcement action. FTC enforcement could entail administrative proceedings or civil litigation seeking a federal judicial injunction and potential monetary penalties. In a significant April 2021 ruling, the Supreme Court [limited the FTC's ability to seek monetary relief](#) in federal court for violations of the FTC Act. Congress needs to clarify that the agency has this statutory authority. The legislation also should make clear that new federal mandates don't preempt application of state consumer protection standards.

As noted, Congress and, in turn, the FTC must frame the new oversight as procedural, not substantive. To mitigate First Amendment complications, legislation should ban the FTC from telling platforms what substantive moderation rules to establish, let alone dictate particular decisions about content. Social media companies would retain full First Amendment protection to determine what they host on their sites. Users' speech would not be subject to government restriction. By forbidding FTC involvement in crafting or applying substantive content standards, Congress can squarely position the agency's oversight as a form of consumer protection, not unlike the FTC's policing of loans with hidden fees or misleading product warranties that don't cover commonly needed repairs.¹¹

This procedural approach would not provide the proverbial silver bullet for harmful content. Platforms would not be obliged to moderate content aggressively. Sites like [Gab](#), [Parler](#), and [Rumble](#), which market themselves as imposing few limits on users and have become havens for right-wing conspiracy theorists, would not necessarily have to change their policies. But they would have to maintain moderation systems sufficient to deliver on the promises they have made to users.

While not a panacea, a requirement of procedural adequacy would nevertheless be meaningful. Social media companies would have to follow through on assurances they make in their community standards and terms of service. They would do this by providing clear rules, easily understood enforcement procedures, and ways for users to appeal moderation decisions. As a practical matter, mainstream social media companies are extremely unlikely to eliminate or even significantly roll back the extensive—but nevertheless inadequate—moderation systems they currently operate. Doing so would lead to inundation by spam, pornography, hate speech, and other unwelcome content that would drive away users and advertisers.

⁹ We borrow the name from the recently introduced [Digital Services Oversight and Safety Act](#). Another bill, the [Algorithmic Accountability Act](#), also would create a digital bureau that would oversee newly mandated self-assessment and disclosure by social media companies and other firms that rely on automated decision making. The latter legislation would reach beyond content moderation to cover a wide range of "critical decisions" concerning health, finance, housing, and education.

¹⁰ According to the advocacy arm of Consumer Reports, the size of the American economy has tripled since 1979, while the [FTC's capacity has decreased 37%](#).

¹¹ The bipartisan [Platform Accountability and Consumer Transparency Act](#) contains useful language on ensuring that the FTC plays a purely procedural role.

FTC oversight would force tech companies to clarify moderation policies and practices that are often opaque and contradictory. Facebook's quasi-independent [Oversight Board](#) repeatedly has urged Facebook to make it easier for users to understand [why posts were taken down \(or left up\)](#) or accounts were [removed](#). At present, this is frequently impossible, in part, because Facebook [keeps many guidelines used in content decisions secret](#).

To mitigate First Amendment complications, legislation should ban the FTC from telling platforms what substantive moderation rules to establish, let alone dictate particular decisions about content.

The new digital bureau should also look into the adequacy of how human content reviewers do their jobs, including how their determinations relate to the far larger body of decisions driven by artificial intelligence. One potentially fruitful line of inquiry would be whether the widespread practice among social media companies of outsourcing human content review diminishes the quality of that crucial work. Based on [research by our Center](#), we have advocated bringing currently outsourced reviewers into social media companies as direct employees who would be more likely to receive effective training, supervision, and mental health counseling, in light of their exposure to traumatizing online material.¹²

Mandated data disclosure: The other main goal of FTC oversight should be the disclosure of currently nonpublic information about how content moderation actually works. This would include how platform algorithms rank, recommend, and remove content, as well as information about the kind of content that reaches large audiences and the factors that cause such content to go viral. These disclosures would help users and advertisers make better-informed choices about which platforms to patronize. Researchers and policymakers would gain insights that point to possible further reform.

Disclosure of data on the influence of algorithms, for instance, could help reveal whether certain harmful content is reaching a wide audience because users are searching for it, platforms are ranking it prominently in user feeds, or recommendation engines are urging particular users to view it. Openness about the mechanics of content amplification might lead to public pressure for platforms to ease their algorithmic drive for user engagement and thereby diminish the spread of hateful, sensationalistic, and patently false content. The FTC-enforced transparency requirements included in the [Digital Services Oversight and Safety Act](#) would go a long way toward achieving these aims.

Congress and, in turn, the FTC should distinguish between disclosure requirements leading to information becoming entirely public versus data which, because of user-privacy concerns or the need to protect proprietary intellectual property, would become available only to regulators and qualified researchers. The [Digital Services Oversight and Safety Act](#) grapples with these distinctions, as does the narrower [Platform Transparency and Accountability Act](#). The latter legislation, which has bipartisan backing, would direct the FTC to require that tech companies provide data to certified outside researchers, subject to certain restrictions protecting user privacy. The FTC would work with the National Science Foundation to determine which inquiries proposed by outside researchers qualify; the companies would have no say. The legislation defines failure to comply with this data-sharing system as "unfair or deceptive" under Section 5(a) of the FTC Act. Disclosure for research purposes would complement the salutary effects of the [Honest Ads Act](#), bipartisan legislation that would make digital political advertising subject to the same transparency obligations now imposed on broadcast political ads.

Lessons from corporate cyber-security cases

Since 2000, the FTC has litigated or settled dozens of enforcement actions against companies that promised customers they would maintain reasonable information-security protections and then failed

¹² The FTC also could mandate disclosure of the size of human moderation workforces and require allocation of sufficient resources to operate moderation systems commensurate with the scale of large platforms.

to do so. Customers whose personal and financial data were exposed by digital breaches were subject to significant potential harm and were not in a position to protect themselves.

A 2015 case involving the Wyndham hotel chain provides an example. The FTC sued Wyndham, alleging that the company's data-security failures led to breaches resulting in millions of dollars of fraudulent charges on consumers' credit and debit cards. Wyndham fought the suit, arguing that the alleged lapses didn't amount to an "unfair" commercial practice. Clearing the case for trial, the U.S. Court of Appeals for the [Third Circuit ruled](#) that the FTC had identified overlapping elements of unfairness and deceptiveness in the hotel chain's conduct: "A company," the court said, "does not act equitably when it publishes a privacy policy to attract customers who are concerned about data privacy, fails to make good on that promise by investing inadequate resources in cybersecurity, exposes its unsuspecting customers to substantial financial injury, and retains the profits of their business."

Analogously, a social media company does not act equitably when it publishes a content moderation policy to attract users concerned about exposure to hate speech, misinformation, and violent incitement; fails to make good on that promise by maintaining a procedurally inadequate moderation system and thereby exposes unsuspecting users to harmful content; and retains the profits derived from digital advertising targeting those users. (In the Wyndham case, the hotel chain ultimately [settled](#), agreeing to maintain a comprehensive information-security program.)

A final word on the First Amendment

Confronted with enhanced oversight by the FTC, deep-pocketed social media companies will file First Amendment court challenges. If the judiciary responds by categorizing the regulation as an exercise of consumer protection, it ought to survive. But if courts see regulation of social media as an indirect attempt to restrict expressive content, they will subject it to "strict scrutiny," which is typically fatal for the government action in question. To survive, a challenged regulation must promote a "compelling government interest" in a "narrowly tailored" manner.

The compelling interest in this case is protecting consumers from unfair and deceptive practices that engender unwanted hateful speech and conduct, political disinformation and erosion of democratic norms, and incitement of political violence. Narrow tailoring is seen in the regulation's emphasis on procedural rather than substantive requirements. Mandatory data disclosure for research purposes is analogous to constitutionally sound data demands the government makes of a vast array of industries in the name of customer safety and product effectiveness.

Social media companies also would attack new regulation as disfavoring them in comparison to traditional news organizations. But this distinction is justified. With some exceptions, *The New York Times* makes a specific decision about each piece of expressive content it places on its website or in its print edition. In contrast, social media platforms are designed so that they can host billions of user-generated posts that appear almost instantaneously and typically without human review. Social media technology created a novel challenge in the potential for harmful content to go viral with dire real-world consequences. Social media companies have responded by building moderation systems, but these systems routinely fall short. This situation necessitates the kind of judicious, non-content-based government oversight we propose.

Section 230

For nearly five years, the congressional debate about regulating social media companies has revolved around [Section 230 of the Communications Decency Act of 1996](#). Revising the 26-year-old provision is not a centerpiece of our policy recommendations, but it deserves mention as an adjunct to enhanced FTC oversight. In brief, we urge Congress to curtail Section 230 in order to incentivize companies to increase their vigilance in moderating forms of hateful content that have proven particularly dangerous.

Section 230 protects providers of “interactive computer services” from being sued over content posted by users or moderation decisions concerning such content. This liability shield has encouraged social media companies to give users wide latitude in what they post without being held legally responsible for the consequences. The second part of the provision provides yet more protection, holding platforms harmless for decisions to take down content or leave it up. Defenders of Section 230 contend, rightly, that it helped the social media industry get aloft in the 2000s and that even today it benefits smaller sites that lack the resources to defend against costly lawsuits. Critics of the law observe, correctly, that its authors couldn’t have anticipated the rise of leviathans like Facebook and Google or the volume of toxic content they spread. These giants, the critics add, now hide behind Section 230 when faced with legitimate allegations that this toxic content has caused real-world damage.

Our view is that Section 230 still fosters free-wheeling expression. It plays an important role shielding platforms—ranging from the user-generated encyclopedia Wikipedia to consumer-review site Yelp—that might otherwise have trouble surviving if confronted by waves of litigation. On the other hand, blanket application of Section 230 to the benefit of Facebook or Google, with their bottomless legal budgets, seems problematic. Legislation should address two goals:

- **Clarify that Section 230 is not a defense against FTC enforcement:** When seeking to hold platforms responsible for protecting consumers by sharing data and/or maintaining procedurally adequate moderation systems, the FTC should not encounter a Section 230 wall.
- **Remove the liability shield for claims related to certain harms:** Several pending bills “carve out” exceptions to Section 230 for suits based on specified types of harmful content.¹³ The [Protecting Americans Against Dangerous Algorithms Act](#), for example, would lift liability protection for suits alleging involvement in foreign or domestic terrorism. It would apply only when such suits point to algorithmic amplification of harmful content, as opposed to chronological ranking or search results. The [SAFE TECH Act](#) would remove the shield for suits involving ads and other paid content, stalking, harassment, and civil rights violations. Congress could adopt all or some of the carve-outs found in these two bills. We agree with the authors of the dangerous algorithms act that carve-outs ought to come into play only in suits alleging that automated systems have amplified harmful content.¹⁴

4. Conclusion

With dozens of competing bills pending in the House and Senate, it isn’t clear which ones will make headway at the committee level or whether congressional leaders will make room on the legislative calendar for floor debate. Still, reining in social media—*somewhat*—is the rare issue on which there is some real momentum on both sides of the aisle. It would be a shame if the profusion of legislative proposals went totally for naught.

Our hope is that the ideas in this paper will help clarify the legislative debate and subsequently shape FTC action—if not this year, then eventually. Online consumer protection is the sort of goal that ought to attract support on both sides of the aisle. But even if the extreme polarization of the moment—exacerbated, not incidentally, by social media—continues to hinder progress in the short run, the need for digital regulation will persist. Sooner or later, we will have to come to grips with the societal toll this alluring technology has taken.

¹³ Several carve-outs to Section 230 are already incorporated into the law: Platforms are legally responsible for civil claims related to federal crimes, sex trafficking, and intellectual property violations.

¹⁴ The [Justice Against Malicious Algorithms Act](#) employs a roughly similar mechanism, removing Section 230 protection when a platform “knowingly or recklessly” promotes harmful content.