Testimony of Professor Nathaniel Persily

James B. McClatchy Professor of Law
Co-Director of the Stanford Cyber Policy Center
Stanford Law School[1]

Before the United States Senate Committee on Homeland Security and Governmental Affairs on "Social Media Platforms and the Amplification of Domestic Extremism and Other Harmful Content"

Submitted October 26, 2021

Thank you, Mr. Chairman and Members of the Committee, for inviting me today to testify on the role of social media in amplifying domestic extremism and harmful content. My name is Nate Persily. I am the James B. McClatchy Professor of Law at Stanford Law School and Co-Director of the Stanford Cyber Policy Center. Perhaps most notably for purposes of this hearing, I was also the cofounder of Social Science One, an effort to get internet platforms, such as Facebook, to share privacy protected data with outside researchers.

I want to use my testimony today to highlight what we know, what we need to know, and then what to do about harmful content and its relationship to social media platform policies.

But before I do that, let me begin my remarks by saying why we are here. We are here because Frances Haugen, the Facebook Whistleblower, produced thousands of pages of internal documents revealing internal research and communication at the company relating to harms Facebook investigated and steps they took or failed to take to combat them. It provided a rare glimpse of the internal workings of the company, and most impressively, the kind of data that informs the platform's assessment of harms and evaluation of potential interventions. Facebook knows an enormous amount about its users and its platform -- and most importantly, Facebook employees are the only ones with access to that information.

That equilibrium – where firm insiders know everything and the rest of us are left to guess – is unsustainable. **Facebook and the other Silicon Valley Platforms have lost their right to secrecy**. We need national transparency legislation that will allow researchers, other than those tied to the profit-maximizing mission of the firms, to get access to the data that will shed light on the most pressing questions related to the effects of social media on society.

[1] Affiliation for identification purposes only; appearing in personal capacity.

I. Harmful Speech Online: What We Know and What We Need to Know

Despite the inability to access data, researchers have learned a lot over the last decade about various online harms.[2] Of course, the harms attributed to social media have multiplied in recent years, as the platforms have been blamed for everything from human trafficking to anorexia to genocide. And as the Haugen revelations depict, even through this year, firm insiders had issued warnings about all of these issues and more.

The 2016 election represented a turning point in the way many people view social media. The disclosure by Facebook, itself, of the efforts of Russian and other foreign agents to meddle in the 2016 campaign quickly turned what had been a utopian view of the potential of social media for democracy to a dystopian one filled with amplified hate speech, disinformation, foreign election interference, and incitement to violence.

Around the same time, because of the Cambridge Analytica scandal, Facebook shut down some of the APIs and other pathways for outsiders to access its data. That scandal, as is well known, involved a researcher at Cambridge University who accessed social graph data and made it available to a political consulting firm. As a result, Facebook paid $5 billion pursuant to a consent decree with the FTC. The scandal casts a shadow over all academic efforts to access platform data.

In the years since the 2016 election and the Cambridge Analytica scandal, researchers have sought to answer the “big” questions about social media’s effect on democracy.[3] I should emphasize at the outset that the study of social media’s effects has been biased considerably toward the United States and Europe, despite the fact that the majority of users of platforms like Facebook now exist elsewhere. We have seen from the Haugen revelations, as well as plenty of earlier reporting, how unprepared Facebook has been in places like India, Myanmar, and Ethiopia. Especially in places where it does not have employees with the requisite language skills, Facebook often cannot enforce its community standards effectively or train classifiers to filter out problematic content. Similarly, although several important studies[4] have been published, we should be hesitant to generalize, particularly from the U.S. experience, as to how social media is affecting democracies around the world. Indeed, as “bad” as things might be here, where a disproportionate share of the “integrity” resources of the firms are directed, it is quite likely that the problems elsewhere are much worse, as the recently disclosed documents suggest.

[2] A disproportionate share of our knowledge is based off of Twitter data, because tweets are public and the platform has been the most open and welcoming of outside researchers. As a result, much of what we know about the internet and social media may be a bit warped by the unique affordances of Twitter, which is much smaller than Facebook or YouTube.

[3] See generally Nathaniel Persily & Joshua A.Tucker, eds., 2020. *Social Media and Democracy: The State of the Field and Prospects for Reform* (Cambridge University Press); Joshua A. Tucker, Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal & Brendan Nyhan, *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (March 19, 2018), https://ssrn.com/abstract=3144139 or http://dx.doi.org/10.2139/ssrn.3144139.

[4] See, e.g., Payal Arora, The Next Billion Users: Digital Life Beyond the West (2019).

A. Amount of Content v. Rates of Exposure

The most difficult challenge in assessing the scale of harmful content online is that no one outside the firms knows how often users are exposed to such content. Whether the category is domestic extremism, terrorist content, election interference, incitement, disinformation or hate speech – those outside the firms do not know the true scale of the problem. Indeed, even those at the platforms have a myopic view because they can only see their slice of social media, and do not know about exposure through other online platforms, let alone cable and broadcast news, talk radio, and other legacy media.[5]

We know there is *a lot* of harmful content online, but there is a lot of all kinds of content online.[6] The key question is who produces, sees, and engages with this content? And how much does the average user, as well as large groups of users at the tails of the distribution, see such content?[7]

For most users of these platforms, their online lives are very similar to their offline lives. The types of people they talk to and the types of information they consume are similar to what they see on television and other media. For many people, in fact, their offline lives lead to exposure to more heterogenous content, given that our weak ties (e.g., high school friends and distant relatives) may be more politically diverse than the people in our immediate vicinity or the news sources we choose to read.[8]

Scholars now realize that, when assessing the scale of harmful online content, focusing on the average user leads to a warped assessment of the problem.[9] Although we

[5] Cf. Jennifer Allen et al., Evaluating the fake news problem at the scale of the information ecosystem, Science Advances, April 3, 2020, https://www.science.org/doi/pdf/10.1126/sciadv.aay3539.

[6] One issue relating to assessing the quantity of online content – harmful or otherwise – concerns the use of automation ("bots") in flooding the information ecosystem. See Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini & Filippo Menczer, The spread of low-credibility content by social bots, Nature Communications (2018), https://www.nature.com/articles/s41467-018-06930-7; Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer, Uncovering Coordinated Networks on Social Media: Methods and Case Studies, Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (2021), https://ojs.aaai.org/index.php/ICWSM/article/view/18075/17878 .

[7] For a sophisticated analysis that tries to grapple with both the "denominator problem" and the definition of hate speech, see Alexandra A. Siegel, Evgenii Nikitin, Pablo Barberá, Joanna Sterling, Bethany Pullen, Richard Bonneau, Jonathan Nagler, & Joshua A. Tucker. "Trumping Hate on Twitter? Online Hate Speech in the 2016 US Election Campaign and its Aftermath." Quarterly Journal of Political Science 16, no. 1 (2021): 71-104.

[8] See Matthew Gentzkow & Jesse M. Shapiro, Ideological Segregation Online and Offline, Quarterly Journal of Economics (2011), Vol. 126, Issue 4, 1799-1839, https://academic.oup.com/qje/article-abstract/126/4/1799/1924154?redirectedFrom=fulltext ; Gregory Eady, Jonathan Nagler, Andy Guess, Jan Zilinsky, &Joshua A. Tucker. "How many people live in political bubbles on social media? Evidence from linked survey and Twitter data." Sage Open 9, no. 1 (2019): 2158244019832705.

[9] See generally Dimitar Nikolov, Alessandro Flammini & Filippo Menczer, Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. HKS Misinformation Review, 1(7). 2021. http://doi.org/10.37016/mr-2020-55.

all have heard stories about relatives who began by joining a knitting club on Facebook and ended up as QAnon adherents, that pathway is less frequented than conspiracy-curious users becoming incrementally more involved with communities defined by their conspiracy of choice or, even more ubiquitous, diehard adherents using the platforms as forums for strengthening their communities and providing a space for organizing.[10]

By all accounts, the producers and consumers of harmful speech – disinformation, hate speech, incitement, etc. – represent a small, but active, dedicated, and sometimes dangerous, share of users.[11] This appears to be true for election[12] and vaccine disinformation,[13] as well as hate speech or extremist racist content.[14] Even a small share of Facebook or YouTube users, though, can still equal millions of people.

This is also why it is difficult to study these problems from the outside. Studies of small shares of radicalized users require very large sample sizes. It is sometimes difficult to recruit people from these communities to be research subjects and random samples might not capture them. The platforms, however, know how large these communities are and the nature of the content they produce and consume. If outsiders had access to the same kinds of data disclosed in the studies from the Haugen revelations,

[10] Shruti Phadke, Mattia Samory, & Tanushree Mitra, What Makes People Join Conspiracy Communities?: Role of Social Factors in Conspiracy Engagement, Proceedings of the ACM on Human-Computer Interaction, Volume 4, Issue CSCW3, December 2020, https://dl.acm.org/doi/abs/10.1145/3432922; Shruti Phadke and Tanushree Mitra, Educators, Solicitors, Flamers, Motivators, Sympathizers: Characterizing Roles in Online Extremist Movements. Proc. ACM Meas. Anal. Comput. Syst. 37, 4, Article 111 (August 2018), https://doi.org/10.1145/1122445.1122456.

[11] See Shruti Phadke & Tanushree Mitra, Many Faced Hate: A Cross Platform Study of Content Framing and Information Sharing by Online Hate Groups, http://faculty.washington.edu/tmitra/public/papers/hategroups-chi2020.pdf .

[12] One study of online election disinformation on Twitter in 2016, for example, finds exposure to be concentrated on a small slice of the population (1%), and sharing of that content occurs among an even tinier slice (0.1%). Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. “Fake news on Twitter during the 2016 US presidential election.” Science 363, no. 6425 (2019), 374-378. See also Guess, Andrew, Jonathan Nagler, and Joshua Tucker. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science advances* 5, no. 1 (2019); Nicolas Berlinski, Margaret Doyle, Andrew M. Guess, Gabrielle Levy, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, and Jason Reifler. 2021. The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections. Journal of Experimental Political Science. https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/effects-of-unsubstantiated-claims-of-voter-fraud-on-confidence-in-elections/9B4CE6DF2F573955071948B9F649DF7A (“[E]xposure to claims of voter fraud reduces confidence in electoral integrity, though not support for democracy itself. . . . Worryingly, corrective messages from mainstream sources do not measurably reduce the damage these accusations inflict. These results suggest that unsubstantiated voter-fraud claims undermine confidence in elections, particularly when the claims are politically congenial, and that their effects cannot easily be mitigated by fact-checking.”).

[13] Andrew M. Guess, Brendan Nyhan, Zachary O'Keeffe, and Jason Reifler. 2020. "The sources and correlates of exposure to vaccine-related (mis)information online." Vaccine 38(49): 7799-7805, https://www.sciencedirect.com/science/article/pii/S0264410X20313116 ; Francesco Pierri, Brea Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, John Bryden, The Impact of online misinformation on U.S. COVID-19 vaccinations, May 2021, https://arxiv.org/abs/2104.10635 .

[14] Annie Y. Chen et al., Exposure to Alternative and Extremist Content on YouTube (2020) at https://www.adl.org/media/15868/download.

we could all better understand the scope of the problem related to harmful speech online and what might be done about it.

B. Algorithmic Delivery v. Search, Subscription, and Selection

How users arrive at harmful content represents a critical question to which the literature is only beginning to provide an answer. Is the most problematic content served to users through algorithms or is it sought out by users interested in the content, who will often subscribe to channels, follow accounts, or become members of groups? The answer, of course, is that users arrive at such content through both paths – algorithmic curation based on past viewing habits, as well as through search, subscription and selection. However, understanding which is a more dominant pathway for radicalization is critical to understanding whether certain types of interventions, such as direct regulation of algorithms, will prove fruitful in reducing the reach of harmful content.

The Facebook Files present damning evidence of internal research describing how the Newsfeed algorithm was delivering harmful content to users around the world. Although Facebook took certain measures to reduce the reach of election-related disinformation, incitement, and hate speech, it still appears their efforts could not catch a large share of the election and COVID-related disinformation flowing over the platform. As several of the Facebook disclosures demonstrate, the machine learning classifiers developed to deal with a lot of harmful content are still in their infancy, and they are better at flagging some community standards violations (such as nudity) than others (such as hate speech).

Understanding the role algorithms play in amplifying harmful content is critical to evaluating the platforms' independent responsibility for the content they themselves are making popular.[15] It has become fashionable to describe Facebook, Google and Twitter as the "new public square." But the online spaces they control are very different than the Boston Commons or a street corner with a soap box. Based on the behavioral history of the user and predictions based on data gathered from everyone on the platform, the algorithms prioritize certain communication over others; they do not allow every speaker to have access to every willing listener at every given time. Although they are not publishers in the traditional sense, the platforms' decisions to prioritize some content over others -- and therefore give it "reach" -- creates greater responsibility than if the platform were merely hosting all comers or prioritized information on a first-come-first-served basis.

The experience with Facebook groups and the platform's recommendations of them demonstrates this dynamic. The most recent revelations tell of how internal researchers created fake accounts to demonstrate that Facebook's algorithm recommended groups and pages that would lead some users down rabbit holes of racist

[15] See Giovanni Luca Ciampaglia, Azadeh Nematzadeh, Filippo Menczer & Alessandro Flammini, How algorithmic popularity bias hinders or promotes quality, Scientific Reports (2018), https://www.nature.com/articles/s41598-018-34203-2.

extremism and conspiracy thinking. Based on these and other internal studies, Facebook banned political group recommendations in the period preceding the 2020 election. But overtly political groups are just the tip of the iceberg; the same dynamics have been seen with respect to anti-vax and QAnon content, as well as other kinds of conspiracies, both political and otherwise. As the disclosures and other independent research has confirmed, these groups have proven especially nimble in adapting to and circumventing the measures platforms use to take them down or demote them.[16]

Perhaps on no other issue is there such a gap between the contentions of the platforms and those of their critics (which includes both whistle blowers and conventional wisdom). Critics say the algorithms play an outsized role in surfacing extremist and sometimes dangerous content.[17] In particular, algorithms that myopically prioritize engagement will keep delivering to users the kind of content they engaged with previously that is likely to keep them on the platform. A user that is curious about conspiracies or other fringe content could become deeply enmeshed in a community dedicated to that issue, the argument goes, if the algorithm keeps recommending fringe pages, channels, or groups. Given that much incendiary and potentially harmful material may generate engagement because of its salacious or emotional appeal (rather than appeal to reason), the algorithm that prioritizes engagement necessarily feeds users more of the problematic content that they have signaled through their watch history would keep them on the platform.[18]

The platforms reject this characterization.[19] They point to measures they take to demote and takedown, rather than amplify, harmful content. They also maintain that they

[16] See generally the work of the Virality Project of the Stanford Internet Observatory at the Stanford Cyber Policy Center, noting the way anti-vax groups visually block key words, include disinformation in comments instead of posts, and share screenshots instead of links to problematic content. See Rachel Moran et al., Content Moderation Avoidance Strategies, July 29, 2021, https://www.viralityproject.org/rapid-response/content-moderation-avoidance-strategies-used-to-promote-vaccine-hesitant-content .

[17] See, e.g., Eslam Hussein, Prerna Juneja, & Tanushree Mitra, Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube, Proc. ACM Hum.-Comput. Interact. 4, CSCW1, Article 48 (May 2020), https://doi.org/10.1145/3392854 .

[18] Much has been written trying to test these hypotheses, especially as they relate to YouTube. See Annie Y. Chen et al., Exposure to Alternative and Extremist Content on YouTube (2020) at https://www.adl.org/media/15868/download ; Manoel Horta Ribiero et al., “Auditing Radicalization Pathways on YouTube,” 2019, https://arxiv.org/abs/7908.08373; Mark Ledwich and Anna Zaitsev, “Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization,” 2019, https://arxiv.org/abs/7912.11211; Kevin Munger and Joseph Phillips, “Right-Wing YouTube: A Supply and Demand Perspective,” The International Journal of Press/Politics, October 21, 2020; Buntain, Cody et al., “YouTube Recommendations and Effects on Sharing Across Online Social Platforms,” ArXiv:2003.00970 [Cs), July 20, 2020, https://arxiv.org/abs/2003.00970; Faddoul, Chaslot, and Farid, “A longitudinal analysis of YouTube's promotion of conspiracy videos.”; Eslam Hussein, Prerna Juneja, and Tanushree Mitra, "Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube," Proceedings of the ACM on Human-Computer Interaction 4, no. CSCW1 (May 28, 2020): 1-27; 24.Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, David M. Rothschild, Markus Mobius, and Duncan J. Watts, "Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube," 2020, https://arxiv.org/ pdf/2011.12843.pdf.

[19] See, e.g., Nick Clegg, You and the Algorithm: It Takes Two to Tango, Mar. 31, 2021 at https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2 ; The YouTube

abandoned a singular focus on raw engagement (which paradoxically led users to spend less time on the platform because engaging content was often low quality) and replaced it with measures of healthy engagement, “Meaningful Social Interactions,” and “Valued Watch Time.” Moreover, they contend that the recommendation algorithms direct people more often to mainstream news and information and, in the case of Facebook Newsfeed, more often toward content produced by friends and families.

Moreover, the existence of conspiracy-mongering, hate speech, and incitement on peer-to-peer messaging platforms complicates the story about the algorithm being the principal source of radicalization. Platforms such as WhatsApp (owned by Facebook) are known, particularly in the developing world, to be founts of the same problematic content found in the Facebook Newsfeed. [20] Yet, WhatsApp has no algorithm and no advertising. Rather, both grassroots and elite influencers use WhatsApp to incite violence, propagate disinformation, and promote hate speech – without the platform even knowing about it since the messages are encrypted and unseeable by the platform.

This debate between the platforms and their critics on the role of recommendations and algorithmic curation in the propagation of dangerous content can only be resolved with outside access to platform data and auditing of the algorithms. The Facebook Files reveal internal researchers’ concerns consistent with those of the critics – namely, that the Newsfeed algorithm and group recommendations amplified COVID and election-related disinformation, content related to the January 6th insurrection, and hate speech in the U.S. and abroad. This debate over the algorithm is also a debate about social media itself – whether a platform can organize information based on user behavioral signals in a way that does not reinforce “unhealthy” choices previously made. Moreover, can they do so in a way that also does not also open them up to claims of censorship and shadow-banning based on ideological or partisan bias?

C. The Poorly Understood Role of Advertising in Propagation of Online Harms

If we array platform “affordances” along a continuum of responsibility, advertising would seem to be the area where platforms have the greatest responsibility for content delivered to users. Whereas the platform is understandably shielded (by section 230 of the CDA) from liability for user-generated content, when the platform takes money for content that it delivers in microtargeted fashion to users calculated to be persuaded by it, the platform’s responsibility for that content is at its apex. Google and Facebook are online advertising monopolies, in the end, so their power over ads above all else requires the greatest scrutiny.

Team, Continuing our work to improve recommendations on YouTube, Jan. 25, 2019, https://blog.youtube/news-and-events/continuing-our-work-to-improve/.

[20] See Ashkan Kazemi, Kiran Garimella, Gautam Kishore Shahi, Devin Gaffney, & Scott A. Hale, Tiplines to Combat Misinformation on Encrypted Platforms: A Case Study of the 2019 Indian Election on WhatsApp, July 23, 2021, https://arxiv.org/abs/2106.04726; Punyajoy Saha, Binny Mathew, Kiran Garimella, Animesh Mukherjee, “Short is the Road that Leads from Fear to Hate”: Fear Speech in Indian WhatsApp Groups, Feb. 7, 2021, https://arxiv.org/abs/2102.03870.

Nevertheless, we know precious little about the role that advertising plays in promoting hate, disinformation, incitement and other illegal activity online. We have examples, most notably from the 2016 election, of Russian agents buying ads on all major platforms to sow division and fan the flames of ethnic, racial, and religious hatred, as well as meddle in the election itself. Moreover, Facebook, at times, has relaxed some of its rules on disinformation and some community standards when it comes to political ads, not wanting to referee "truth in advertising" during an election campaign. However, one non-peer reviewed study found that when Facebook initiated a ban on advertising from fake news websites following the 2016 election that sharing of fake news decreased substantially.[21]

To understand the relationship of advertising to dissemination of harmful content, it is important to recognize the differences between online and legacy media advertising. It may be more appropriate to consider not advertising per se, but paid versus organic content and the interaction between the two. To be sure, we have many examples of inciting and dangerous content purchased through traditional advertising. Indeed, the audacity of the Russian intervention in 2016 can be seen in the exploitation of the traditional ad platforms to send simple, if polarizing, messages (sometimes purchased with rubles) to large numbers of users, as well as organize events and recruit followers.

Beyond traditional and familiar forms of advertising, though, amplification, itself, is for sale by the platforms. Traditional media entities realize this, which is why many of the most prominent legacy publications pay to boost their reporting on both Twitter, Google, and Facebook. However, these same opportunities are available to fringe groups, foreign actors, individuals, and truly fake news sites. The platforms identify content as promoted or sponsored (sometimes in small type), but it is very difficult to distinguish advertisements from organic content. Indeed, sometimes the posts can be identical -- the same story might arrive into a user's newsfeed because a friend has forwarded it or because the entity behind it paid to place it there. The distinctive feature about all paid content on modern social media platforms, though, is the microtargeting that the platforms make for sale. Therefore, not only can a publication (or bad actor) amplify content for a broader audience, they can also target a narrow one either by providing their own lists of individuals to create a custom audience or using the tools the platforms make available to target based on a myriad of characteristics.

Because Russian advertising efforts earned such infamy after 2016, the platforms developed political ad libraries to provide for greater transparency.[22] In addition, they adopted verification procedures for political ads (involving postcards sent to actual domestic addresses), to ensure election-related advertisements are not purchased by

[21] Lesley Chiou & Catherine Tucker, 2018. "Fake News and Advertising on Social Media: A Study of the Anti-Vaccination Movement," NBER Working Papers 25223, https://ideas.repec.org/p/nbr/nberwo/25223.html

[22] See Erika Franklin Fowler, Michael M. Franz, & Travis N. Ridout, Online Political Advertising in the United States, in Social Media and Democracy: The State of the Field and Prospects for Reform Ch. 6 (Persily, N. and J. Tucker eds., Cambridge University Press, 2020).

foreign actors. For the 2020 election, moreover, Twitter banned election ads and Facebook banned new ads for the last week of the campaign.

Outside efforts to evaluate the completeness of the ad libraries have run into problems, however. Most notoriously, Facebook this past summer suspended the accounts of researchers at NYU who were studying political ads.[23] The researchers developed a browser plug-in (the Ad Observer) that users could install to scrape the advertising information they were seeing when logged into Facebook and other sites. They discovered that many political ads never made it into the ad library[24] and that ads violating Facebook ad policies would find ways to avoid detection.[25] The NYU researchers remain suspended, even though the Federal Trade Commission took the extraordinary step of issuing a letter saying the plug-in would not run afoul of the FTC's privacy-related consent decree with Facebook. As it now stands, outsiders are severely handicapped in their efforts to evaluate the relationship of advertising and other boosted content to a variety of online harms.

II. What to do about it? Regulation and Transparency Relating to Harmful Content and Social Media

The fiasco involving the NYU Ad Observatory, as well as my own experience trying to facilitate outsider access to social media data with Social Science One, has convinced me that only federal legislation will open up these platforms to outside scrutiny. Researcher access to platform data (discussed in the Appendix) is only one aspect of the necessary transparency, and transparency is only one component of a larger legislative agenda relating to harmful online content. Transparency legislation may be the constitutionally safest form of regulation, though.

To level set, most direct regulation of harmful, but legal, online content would violate the First Amendment. With a few notable exceptions,[26] the regulation of hate speech, disinformation and (most forms of) incitement, cannot be done through outright bans or takedown mandates. For all the talk about repealing or amending Section 230 of the Communications Decency Act, no change in that law will get at most of the problems discussed here. It might expose the platforms (and potentially all platforms, including small startups trying to challenge established players) to liability for defamation and some examples where the connection between online speech and offline violence is most clear. But most forms of election or COVID-related disinformation, let alone hate speech

[23] Laura Edelson & Damon McCoy, We Research Misinformation on Facebook. It Just Disabled Our Accounts, NY Times, Aug. 10, 2021 (noting that their study suggested 100,000 ads that were not included in the ad library).
[24] See Tony Romm & Isaac Stanley-Becker, Tens of thousands of political ads on Facebook lacked key details about who paid for them, new report finds, Washington Post, Mar. 8, 2020.
[25] Jeff Horwitz, Political Groups Elude Facebook's Election Controls, Repost False Ads, Wall Street Journal, Nov. 1, 2020.
[26] Those exceptions might concern foreign agents intervening in elections, as well as some forms of child protection.

and incitement that does not satisfy the high standard the Court has set for direct advocacy of (likely) violence, cannot be regulated outright.

All of that said, there is still a lot that Congress can do. A case in point – certain forms of regulation of online advertising, in general, and microtargeting, in particular, should pass constitutional muster. Taxation of online advertising revenue might hit these firms the hardest. Even transparency (that is, compelled disclosure) as to purchaser identity, ad content, targeting, and exposure would help prevent the purchased amplification for harmful content. These measures should be applied to all ads, not just political ones.

Second, content neutral forms of regulation such as privacy and competition (antitrust) protections should not face high First Amendment hurdles. It might seem that these types of interventions, while necessary to serve other interests, are unrelated to the propagation of online harmful content. But restricting the kinds of data that large platforms collect about their users is one more way to inhibit microtargeting of ads and other content. Moreover, antitrust scrutiny, even if it falls short of breaking up these firms or treating them like quasi-public utilities, might lead to measures that allow for entry of new social media companies. In particular, if data portability and interoperability become part of the antitrust agenda,[27] the power of Google and Facebook over the information ecosystem (as well as the importance of their community standards and their enforcement) will be muted.[28]

Finally, the transparency agenda must be broad and multifaceted. It should include the kind of researcher access described in the Platform Transparency and Accountability Act below. But it should also require production of data in publicly available (but privacy-protected) interfaces that would allow journalists and the general public to understand in broad terms user exposure and engagement with content.[29] Similarly, algorithms for newsfeeds and recommendation engines must be subject to independent audits to unearth hidden biases and vulnerabilities. Algorithmic transparency might be too much to ask for, since these algorithms change almost daily and public disclosure of the algorithm may enable bad actors to game it. Moreover, there is almost no single person at the firms who understands the code, which has been built and rebuilt for the last twenty years or more. But just as we should be able to measure

[27] See Francis Fukuyama, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed, Marietje Schaake, Report of the Working Group on Platform Scale 2020, https://pacscenter.stanford.edu/wp-content/uploads/2020/11/platform_scale_whitepaper_-cpc-pacs.pdf.

[28] It should be noted, however, that having a dozen smaller Facebooks and YouTubes is not necessarily an easier environment from a security perspective. Those two umbrella firms are able to spend untold sums on integrity and security because they are bundled together. So, Instagram and WhatsApp, for example, freeride off of the security teams at Facebook and would not have as developed security if they were to fend for themselves. Similarly, moves for portability and interoperability must reconcile with the privacy tradeoffs inherent in any system that either allows one to "take one's data" to a different platform or forces platforms to make their data treasure troves available to a set of outside competitors.

[29] I have in mind here richer versions of products like Google Trends and Crowdtangle, which would provide aggregated data that extends beyond mere engagement to include actual exposure.

emissions from a car’s tailpipe, we should be able to evaluate exactly what the algorithms spit out under different conditions.

The revelations contained in the Facebook Files represent a turning point for our understanding of one of the most powerful corporations ever to exist. The picture they paint is one of an institution that is simply incapable of managing the technology that it has unleashed on the world. Even with the best of intentions and proper allocation of resources by these powerful firms, though, the “social” aspect of social media will inevitably create harms that elite actors cannot contain. Social media in an unhealthy and polarized society will reflect the underlying fissures that are tearing communities apart. This is why outsiders’ understanding of what is happening online is so important. All stakeholders – civil society, governments, the firms themselves and users – have a role to play in counteracting the downsides and preserving the upside of this new technology. Only if some outside, independent entity untethered to the profit maximizing mission of the firm can regularly access and interpret the data revealed in these new disclosures can we come to grips with the measures that might be necessary prevent the harms to users and society that the Facebook Files have revealed.

A Proposal for Researcher Access to Platform Data: The Platform Transparency and Accountability Act

Nathaniel Persily

The disclosures of whistleblower Frances Haugen have provided a unique glimpse into Facebook's internal research and the ways that the company evaluates and addresses different harms on the platform. As explosive as the content contained in Haugen's revelations may have been, most of the reaction may have arisen from the mere fact that outsiders got an opportunity to see what Facebook knows (or could know) about its users and the information ecosystem it controls. Every inadvertent disclosure that comes out of Facebook gains such notoriety because most of what the public normally sees is subjected to rigorous vetting, corporate-speak and spin.

We should not need to wait for whistleblowers to blow their whistles, however, before we can understand what is actually happening on these extremely powerful digital platforms. Congress needs to act immediately to ensure that a steady stream of rigorous research reaches the public on the most pressing issues concerning digital technology. No one trusts the representations made by the platforms themselves, though, given their conflict of interest and understandable caution in releasing information that might spook shareholders. We need to develop an unprecedented system of corporate data-sharing, mandated by government for independent research in the public interest.

This is easier said than done. Not only do the details matter, they are the only thing that matters. It is all well and good to call for "transparency" or "data sharing," as an uncountable number of academics have, but the way government might set up this unprecedented regime will determine whether it can serve the grandiose purposes tech critics hope it will.

As with so many areas of tech regulation, transparency laws come with tradeoffs. In some cases, for instance, transparency might inhibit necessary security or harm prevention measures, as public disclosures about platform standards' enforcement might lead to gamesmanship by bad actors. When it comes to data access for research, the chief risk that needs to be addressed is user privacy. The shadow of Cambridge Analytica is cast over any academic access to user data, as that scandal involved a university researcher mishandling user data for the benefit of a private political consulting firm. If user data cannot be protected, then the public will not have faith in any government-mandated data-sharing program.

It is critical to understand at the outset, though, that user data is already collected and analyzed – but only by employees at the firms themselves. The threshold question when it comes to outside researcher access is whether the firms (and their employees who are tied to their profit maximizing mission) should have a monopoly on the insights that access to such data guarantees. Perhaps the firms should be prevented from gathering so much user data, but once they do, the public needs to be aware of it and to benefit from the insights that independent analysis will provide.

These benefits will be substantial. Most importantly, the mere fact that outsiders will have access to platform data will affect platform policies and behavior. Digital platforms, like any other association, institution or individual, will alter their behavior if they know they are being watched. Second, researcher access will enable evaluation and auditing of platform rules and interventions to gauge the responsibility of firms for problems that occur on their platforms. In other words, researcher access can enable outside auditing of actions taken by platform against users and

doi:10.1017/xxxxxxx xxxx

content. Third, such access will inform policy makers seeking to regulate the platforms: only if they understand what is actually going on online might they be able to craft the appropriate regulations related to antitrust, privacy, advertising, child safety, content moderation or anything else. Finally, research on digital trace data is absolutely critical to understanding the sociology of the online information ecosystem, irrespective of potential links to policy. A large share of the human experience is taking place online. To understand it we need access to the relevant data.

The proposed legislation that follows – the Platform Transparency and Accountability Act – intends to design a data sharing program that protects user privacy to the extent possible while ensuring outside independent research on platform data. There are many ways to craft such a regime, and I hope this proposal sparks alternative approaches. The key features of any such system, though, must be (1) access by researchers not chosen by the firm to (2) the same data that the firms' own data analysts can analyze but (3) in a secure environment that minimizes any risks of disclosure of user private data.

Any proposal for outside access to platform data must wrestle with several questions (and this list is necessarily underinclusive). First, to which companies or platforms should such a regulatory regime apply? Second, who should have access? Third, to what data should they have access? Fourth and most important, how shall such access be regulated to protect both user privacy and research integrity?

1 Which Platforms?

Google and Facebook are first among (un)equals when it comes to the sheer volume of social media and digital trace data the firms possess. Any regulatory regime aimed at researcher access should be reverse engineered to capture those two firms in particular, as well as TikTok, which is quickly becoming a real competitor to YouTube. Twitter, which already provides more data than any other firm for researcher access, could also be added to the list, if the focus of the regulation is social media, per se.

But what about Amazon, Apple, and Microsoft? Researchers could gain enormous insight from access to those firms' data. Amazon, in particular, represents a monopoly of a different sort with data on users that could be extremely helpful to understanding the digital economy. Moreover, if the communications ecosystem is the target for research, what about the cable and cell phone companies, such as Comcast and Verizon? Surely, they possess data farther down the stack that could be helpful in assessing some relevant problems. A similar argument could be made for traditional media companies, e.g., Fox, or "new media" companies, such as Netflix.

To some extent, the universe of firms to which a data access regime would be applicable depends on the range of phenomena one considers worthy of study and the inability of researchers to gain insights from the outside. For those (like me) for whom the principal concern is the health of the information ecosystem and its impact on democracy, Google, Facebook, and Twitter reign supreme. The identification of the relevant firms, then, would include a definition of social media or search firms meeting some threshold of daily or monthly active users.

The Honest Ads Act[1] took a stab at such a definition in its attempt to force a disclosure regime on online political advertising. That bill defined an "online platform" as "any public-facing website, web application, or digital application (including a social network, ad network, or search engine) which...has 50,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months." The legislative proposal that follows here lowers the bar to 40,000,000 monthly active users in order to capture TikTok as well.

[1] https://www.scribd.com/document/409188376/Mcg-19321

2 Which Researchers?

Deciding which researchers shall have access is one of the biggest challenges to legislation in this area. “Researchers” come in many forms and a wide variety of civil society actors have an interest in the data held by internet platforms. However, some quality control must exist lest political operatives and propagandists repurpose themselves as “researchers” to gain access to platform data. It may also be that a separate regime for platform data access could be erected for think tanks or journalists, many of whom (such as Pew, ProPublica, the Markup, Buzzfeed or the Guardian) have done foundational research on these types of topics. Although categories such as journalists or think tanks may be difficult to cabin and enforce, transparency legislation should have as its goal making as much information available to as many watchdog groups, consistent with the privacy interests of users.

Focusing a data access regime on university-affiliated researchers has several advantages, however. First, a university is an identifiable “thing,” and while low quality academic institutions exist, regulations can more easily specify the type of institutions that house the academics that should be granted access. Second, universities can be signatories to data access agreements with the platforms so as to add another layer of security (and retribution) against researcher malfeasance. Third, universities have Institutional Review Boards (IRBs) that can provide ethics and Human Subjects review for research proposals. Admittedly, IRBs have many well-known problems, but they are existing institutions that are in the business of evaluating research projects and the implications for human subjects. Fourth, in the wake of the Cambridge Analytica scandal, which involved an academic operating outside of his academic capacity, involving universities directly in the process of vetting and vouching for their researchers will make clear to the platforms which researchers are nested in a larger regulatory, contractual, and employment framework. Fifth, the National Science Foundation, which would play a role in vetting researchers, has established procedures in place to vet research projects and researchers from universities.

3 What Data?

In some settings, it is quite easy to define the data that should be made available for research. For instance, when drug trial data are made available for outside review, there are settled and familiar expectations for what kind of information the pharmaceutical company will provide. For Google and Facebook, though, the volume and variety of data they possess are so vast that any legally defined data access regime cannot simply say “turn over all available data to researchers.” Some kind of principle should specify the range of data that should be available for research, or at least a process for deciding what data should be made available.

At a minimum, researchers should be allowed to analyze any data that is otherwise for sale to commercial entities or advertisers. If the datasets are available for a price, then they can be made available for academic analysis. Similarly, any data that goes into the preparation of government or other reports, such as those relating to enforcement of community standards (e.g., how many pieces of content were designated as hate speech and taken down) should be made available.

Beyond that, the key types of datasets that should be made available relate to “who” viewed/engaged with “what” content “when” and “how.” In other words, to answer the most pressing questions relating to social media, we need data that can assess which types of people (though not individuals themselves) were seeing certain online content at certain times. The platforms already collect data of that nature. As part of the regulatory process, the platforms should be forced to identify datasets already in their possession, as well as data that are regularly collected. Then, the FTC, working with the NSF, should establish an application process for projects targeting those datasets. In addition, in order to prevent platforms from suddenly changing their data retention practices now that they are subject to oversight, the enforcement authority (here, the FTC) should have the authority to require the production of datasets deemed reasonably necessary for providing answers to questions researchers ask.

Moreover, the FTC should require the platforms to produce the code necessary to describe how the data were gathered and assembled, and to describe the chain of custody of the dataset. Researchers need to understand how the platform came up with the dataset. The platforms should also be fined if they misrepresent the origins of the data or otherwise produce a dataset inconsistent with what was requested.

All such data must be anonymized or pseudonomized. Moreover, if it can be done without degrading the quality of research, technologies such as differential privacy or the construction of synthetic datasets should be encouraged. In other words, user data must be presented in a format that protects user privacy as much as possible while maintaining utility for the research project.

4 How Shall the Data be Analyzed While Protecting User Privacy?

One of the reasons that the legislative proposal presented here vests enforcement authority in the FTC is that the FTC has been on the frontlines of enforcing privacy promises (to the extent that it is authorized to do so). The consent decree with Facebook following the Cambridge Analytica scandal, for which Facebook was required to pay a $5 billion fine, was negotiated and enforced by the FTC. In an ideal world, the United States, like Europe, might have a cabinet level position that is responsible for digital services, but if any progress on researcher access is to be made in the next two years, it will need to work with existing agencies. The FTC, working with the National Science Foundation, is the logical choice. That agency, then, will be responsible for vetting researchers and research projects and specifying the conditions under which research shall be conducted.

Although the government will be heavily involved in enforcing the program of researcher access, the datasets themselves should never be placed in government hands. It is absolutely critical that there be no risk of government surveillance or privacy intrusions as a result of this program. Alternative models of access would place the datasets in a government-controlled researcher sandbox, which would allow the government to control directly the environment in which data are analyzed. Doing so would necessarily run the risk that at some point in the future, government officials would see this research environment as a honey pot for intelligence and law enforcement activities.

Under the proposal that follows, the data reside at the firm, which is responsible for maintaining security of the research environment and monitoring all research conducted therein. Researchers need to be monitored whenever they are in touch with the data. Every keystroke must be recorded as the data analysis is conducted. Researchers may not take any data out of the research environment without a privacy review being conducted. That includes immediately prior to publication – all publication drafts must be given a privacy review to ensure no data leakage. And in the event that a researcher engages in malfeasance both the researcher and the affiliated university shall be legally liable (even criminally liable) for any privacy violation. We need to make sure measures are in place that reassure the public that no individual's data is of interest to the research project, just the aggregated findings derived from them.

If the platform follows all applicable regulations concerning protecting privacy in the research environment, then it will be immune from suit for the fact that it made such data available under this program. To be clear, this does not immunize them from harms identified by the researchers. If the platform is discovered to be acting fraudulently or contributing to offline harm, then that information might later end up in a lawsuit or even a criminal prosecution. The point about legal immunity here is that the platforms cannot simultaneously be forced by the law to provide data to researchers and then be subject, for example, to a state tort law claim for violations of privacy.

5 Conclusion

Researcher access is only one component of transparency regulation, and transparency legislation is only one component of tech regulation. Nothing in this proposal should be seen as preventing broader reporting obligations for the platforms or construction of public facing APIs. Indeed, we should strive for a system in which any data on issues of public concern relating to the online information ecosystem should be available to the public, if it can be done in a privacy-protective way without other security risks.

One provision in the proposed legislation goes in that direction by dealing with the problem of scraping data from public-facing platforms. It would shield researchers from criminal or civil liability for scraping of public data from large platforms, like Facebook and YouTube. Of course, people disagree about what data, in fact, are "public" on these platforms. However, for researchers who scrape, they cannot be subject to money damages or criminal liability. This would not solve the problem faced by the NYU Ad Observatory, which had its accounts taken down by Facebook since it promoted a plug-in that allowed users to scrape their Facebook. But it would shield them from further actions, such as lawsuits that the platforms might initiate to get damages for terms of service violations arising from scraping.

A similar impulse underlies "Aaron's Law"[2] introduced by Representative Zoe Lofgren and Senator Ron Wyden. In a now famous and tragic episode, Aaron Swartz downloaded a large number of articles from the digital repository, JSTOR. In doing so, he breached the applicable terms of service for the website. Swartz was later arrested and prosecuted under the CFAA, which could have led to a penalty of 35 years in prison and up to $1 million in fines. However, he committed suicide before he was brought to trial. Aaron's Law would remove the threat of a felony prosecution for breaching terms of service in actions like this, if they do not cause significant economic or physical damage.

Just as researcher access is not coterminous with transparency, transparency does not address all problems that tech regulation seeks to solve. Nothing in this proposal should be seen as taking the place of proposals to address competition and antitrust, child safety, advertising, content moderation, cybersecurity and privacy. Indeed, a proposal like the one that follows should be bundled together with federal privacy legislation or other broad regulations of the tech industry.

Researcher access, however, is a condition precedent to effective tech regulation. Right now, we do not know what we do not know. There are fundamental inconsistencies between platform's public representations and those made by whistleblowers, let alone those that feed conventional wisdom. For example, on the critical question of whether algorithms and recommendation systems are leading users toward extremism or promoting disinformation, the defenders and critics of platforms fundamentally degree on basic facts. Policy makers need and deserve answers to these kinds of questions. Only if the government develops and mandates outside researcher access might we be able to get the answers necessary to make effective policy. Otherwise, we will be left with whatever studies the platforms choose to release or whatever research whistleblowers take with them on the way out the door.

Author

Nathaniel Persily is the James B. McClatchy Professor of Law at Stanford University and Co-director of the Stanford Cyber Policy Center.

[2] https://www.congress.gov/bill/113th-congress/senate-bill/1196

Conflict of Interest

Not applicable.

Nate Persily
(in private capacity)

An Act

To support research about the impact of digital communication platforms on society by providing privacy-protected, secure pathways for independent research on data held by large internet companies.

SEC. 1. Short Title

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, that this Act may be cited as the "Platform Transparency and Accountability Act."

SEC. 2. Congressional Findings and Purpose

(a) The Congress finds that certain of the Nation's largest internet platforms exert unprecedented control over the speech marketplace.

(b) Exploitation of the affordances of these platforms has threatened the safety and integrity of our electoral processes, has increased our vulnerability to propaganda attacks by hostile nation-states and domestic extremists, has led to promotion of off-line violence, and has misled the public as to critical facts necessary to promote public health and well-being.

(c) Because of the unprecedented control these platforms exercise over massive amounts of user data and the speech marketplace, Congress finds it necessary to promote independent research on those platforms in order to reveal and help address societal and individual harms caused or exacerbated by these new technologies.

(d) The Congress declares it to be its purpose and policy, through the exercise of its powers to regulate commerce among the several States and with foreign nations and to provide for the general welfare, to assure so far as possible free and fair elections in this Nation so as to preserve our republican form of government, guard against foreign propaganda, and ensure the free flow of information in interstate commerce –

(1) by providing that Qualified Platforms and Qualified Researchers have separate but dependent responsibilities, interests, and rights with respect to obtaining data and information for Qualified Research Projects that will benefit the public good while protecting the privacy rights of the individual user;

(2) by authorizing the Federal Trade Commission to set mandatory data and information sharing requirements applicable to Qualified Platforms affecting interstate commerce, and by creating a Platform Transparency and Accountability Division for carrying out adjudicatory functions under the Act;

(3) by providing the groundwork for understanding the prevalence and character of disinformation, hate speech, and harmful and illegal content spreading by way of large Qualified Platforms, as well as potential political bias in content moderation practices and in algorithmic prioritization of content on those platforms;

(4) by investigating the exploitation of platform affordances by domestic and foreign actors seeking to undermine United States democracy and confidence in the election infrastructure;

(5) to ensure that the market power of certain Qualified Platforms does not pose anticompetitive effects that restrain the information economy; and

(6) to help inform policy makers and regulatory agencies by promoting an accurate understanding of the practices of Qualified Platforms and the dynamics of social media.

SEC. 3. Definitions

For the purposes of this Act --

(1) The term "commerce" means trade, traffic, commerce, transportation, transmission, or communication among the several States, or between any foreign country and any State, or between any State and any place outside thereof.

(2) The term "Commission" or “FTC” means the Federal Trade Commission established under the FTC Act.

(3) The term “Chair” means the Chair of the Federal Trade Commission.

(4) The term “Director” means the Director of the Platform Transparency and Accountability Division appointed by the Chair of the Federal Trade Commission.

(5) The term “Division” means the Platform Transparency and Accountability Division within the Federal Trade Commission.

(6) The term “Personal Information” means any information that is reasonably capable of being associated with a particular individual.

(7) The term “Qualified Platform” means a large, consumer-facing, online or internet-accessible business that meets the criteria for the same established by the Division or its appropriate delegate. Qualified Platforms must have over forty million active monthly Users of their service in the United States and shall include, but are not limited to, any provider of a large online platform, including an online social media service, which, at the request of a recipient of the service, stores and disseminates information to the public.

(8) The term “Qualified Data and Information” means information from a Qualified Platform that meets the criteria for the same established by the Division or its appropriate delegate. Qualified Data and Information may include information about User exposure, engagement, and other behaviors; data about content producers and content production policies; information that the Qualified Platform otherwise makes available for sale to commercial entities or advertisers; information that goes into the preparation of reports that Qualified Platforms provide to the government or other entities, such as those relating to enforcement of community standards; and metadata related to any of the preceding categories.

(9) The term “Qualified Researcher” means a university-affiliated researcher conducting research according to a research plan that has been approved by the Division or its appropriate delegate. No employee of a state of federal law enforcement agency or any government employee except for a university-affiliated researcher shall be considered a Qualified Researcher.

(10) The term “Qualified Research Project” means a research plan that has been approved by the Division or its appropriate delegate.

(11) The term "State" includes any state within the United States, as well as the District of Columbia, Puerto Rico, the Virgin Islands, American Samoa, and Guam.

(12) The term "Scrape" or “Scraping” refers to the act of electronically collecting data that Platforms make available via the user interface, either manually or through an automated process.

(13) The term “User” means a person or entity that uses a social media platform or online marketplace for any purpose, including advertisers and sellers, regardless of whether that person has an account or is otherwise registered with the platform.

SEC. 4. Obligations and Immunity for Qualified Platforms

(a) Each Qualified Platform shall comply with applicable federal, state, and local information sharing and privacy laws and regulations as well as all rules, standards, regulations, and orders issued by the FTC pursuant to this Act which are applicable to their own actions and conduct.

(b) In order to meet its obligations under this Act, a Qualified Platform must provide reasonable privacy and cybersecurity safeguards for the Qualified Data and Information that the Platform shares with Qualified Researchers. Such safeguards, at minimum, shall include

(1) encryption of the data in transit and at rest;

(2) delivery of data in a format determined by the Division that is not reasonably capable of being associated or linked with a particular individual;

(3) use and monitoring of a secure environment to facilitate delivery of the Qualified Data and Information to Qualified Researchers while protecting against unauthorized use of such data;

(4) evaluation by the Qualified Platform of any results garnered by Qualified Researchers before submission for publication but only to prevent public release of Personal Information or other violations of law.

(c) No cause of action under state or federal law relating to or arising solely from the release of data to Qualified Researchers may be brought against any Qualified Platform that complies with this Act and the privacy and cybersecurity provisions described herein.

(d) The legal immunity provided by subsection (c) shall extend only to the fact that data was made accessible to outside researchers and shall not extend to liabilities arising from findings discovered as a result of such research.

SEC. 5. Obligations and Immunity for Qualified Researchers

(a) Qualified Researchers shall be actively engaged in conducting research under a research plan, which was approved by the Division or its appropriate delegate based on its assessment of (1) the intellectual merit of the project (i.e. its potential to advance understanding the impact of large digital communication platforms on society); and (2) its broader impacts (i.e. the project's benefit to society).

(b) Each Qualified Researcher shall comply with

(1) applicable federal, state, and local information sharing and privacy laws and regulations as well as all rules, standards, regulations, and orders issued by the FTC pursuant to this Act which are applicable to their own actions and conduct; and

(2) a prohibition on any attempt to reidentify, access, or publish Personal Information based on Qualified Data and Information that a Qualified Researcher may receive.

(c) No cause of action arising solely from Qualified Researchers' access and use of Qualified Data and Information may be brought against Qualified Researchers who conduct Qualified Research Projects in compliance with this Act and abide by all information sharing and privacy standards described in (a). This immunity includes immunity from potential liability under applicable federal, state, and local laws, as well as any potential liability for a violation of a Platform's Terms of Service that arises solely from the Qualified Researchers' access and use of Qualified Data and Information.

SEC. 6. Sharing of Qualified Data and Information by Qualified Platforms

(a) The Commission shall prescribe regulations requiring that Qualified Platforms maintain and provide Qualified Researchers access to Qualified Data and Information and accurate records of Users' interactions with or exposure to Qualified Data and Information.

(b) Qualified Platforms will be required to provide Qualified Researchers access to Qualified Data and Information as prescribed by regulation;

(c) Qualified Platforms will be required to provide a data codebook that outlines the structure, contents, and layout of the data and must provide the Qualified Researchers with methodological details on how data were collected, cleaned, or manipulated.

(d) Qualified Platforms must enable Qualified Researchers to preserve access to Qualified Data and Information as necessary to carry out and replicate Qualified Research Projects.

(e) The Commission shall also issue regulations requiring that Qualified Platforms, through posting of notices or other appropriate means, keep Users informed of their privacy protections and the information that the Qualified Platform is required to share with Qualified Researchers under this Act.

(f) Any Qualified Data and Information obtained or provided under this Act shall be obtained with a minimum burden upon the Qualified Platform, although Qualified Platforms may not shift the cost of their compliance with this Act to the Qualified Researchers. Unnecessary duplication of efforts in obtaining information shall be reduced to the extent feasible.

(g) Qualified Researchers are authorized to compile and analyze Qualified Data and Information under this section and are required to make all reports created from that analysis freely available to the public in both summary and detailed form.

(h) Twenty (20) business days prior to public release of an analysis by a Qualified Researcher based on Qualified Data and Information or at a time designated by the Division, the Qualified Researcher shall submit a pre-publication version of their research to the Qualified Platforms and the Division for evaluation to confirm that the analysis does not expose Personal Information.

(1) Qualified Platforms may object to the publication or release of any analysis that will necessarily expose Personal Information or otherwise violate federal, state, and local information sharing and privacy laws and regulations or any applicable rules, standards, regulations, and orders issued by the FTC. Such objections must be made in writing to the Division or its delegate within ten (10) business days of the date that the Qualified Researcher submitted the pre-publication version of the research or at a time specified by the Division. Such objections shall include proposed changes to the publication to address the legal problems identified.

(2) If no objection is timely made by a Qualified Platform or the Division, the research may be published.

(3) If objection is timely made by a Qualified Platform, the Qualified Researcher will have ten (10) business days to modify the publication and re-submit it to the Division, which shall decide within ten (10) business days whether the publication complies with the Act. If the Division finds that the publication does not comply with the Act, the Qualified Researcher may appeal such finding to the U.S. Court of Appeals for the Federal Circuit.

(i) The Commission shall have the authority to make, amend, and rescind, in the manner prescribed by 5 U.S.C. § 553, such rules and regulations as it may deem necessary to carry out its responsibilities under this Act.

(j) Access to Qualified Data and Information shall not be granted to any Qualified Research Project pursuant to this statute if it has not been approved or deemed exempt by an Institutional Review Board at the researcher's affiliated university.

SEC. 7. Scraping of Data from Qualified Platforms for University-Affiliated Research

(a) Any university-affiliated researcher conducting research that has been approved or deemed exempt by an Institutional Review Board at the researcher's affiliated university shall be immune from civil or criminal liability for the scraping of data made available through the user interface on Qualified Platforms, regardless of whether the researcher is a Qualified Researcher or the research project is a Qualified Research Project.

(b) Researchers who meet the requirements in (a) are not required to notify Qualified Platforms about Scraping practices, and no cause of action may be brought against Qualified Researchers arising from Scraping practices that comply with this section and such conduct shall be deemed authorized conduct for purposes of the Computer Fraud and Abuse Act (CFAA), 18 U.S.C. 1030.

SEC. 8. Platform Transparency and Accountability Division

(a) It is the purpose of this section to enable and approve Qualified Researchers to carry out the types of Qualified Research Projects set forth in this Act.

(b) There is hereby established within the Commission a Platform Transparency and Accountability Division. The Division shall be headed by a Director who shall be appointed by the Commission Chair, with the approval of a majority of the Commissioners, and who shall serve for a term of four years unless previously removed by the Chair.

(c) The Division is authorized to develop and establish recommended standards, criteria, and approval process for Qualified Researchers, Qualified Research Projects, Qualified Data and Information, and Qualified Platforms under the processes for notice and comment rulemaking in 5 U.S.C. § 553.

(d) The Division shall publish within six months of enactment of this Act and thereafter as needed but at least annually a list of its criteria for identifying Qualified Researchers, Qualified Research Projects, Qualified Data and Information, and Qualified Platforms. Qualified Researchers may suggest platforms for inclusion. Criteria for qualified researchers shall not include consideration of political views, race, gender, gender identity, ethnicity, sexual orientation, age, or disability, although they may express preference for projects proposed by residents of the United States. No person may be qualified as a Qualified Researcher if they act as an Agent of a foreign power as defined by 50 U.S.C. § 1801.

(e) The Division is authorized to inspect data and question Qualified Platforms about the Qualified Data and Information they are making available to the Commission and to Qualified Researchers.

(f) The Director may issue formal written guidance to persons subject to the Act, provided that the Director shall publish all such guidance within six months of its issuance, with the names of the parties and any trade secret or other confidential information redacted.

(g) In addition to any authority vested in the Division by other provisions of this section, the Director, in carrying out the functions of the Division, is authorized to

(1) prescribe such regulations as the Director deems necessary governing the manner in which its functions shall be carried out;

(2) convene an advisory board from relevant Qualified Platforms and Qualified Researchers;

(3) receive money and other property donated, bequeathed, or devised, without condition or restriction other than that it be used for the purposes of the Division and to use, sell, or otherwise dispose of such property for the purpose of carrying out its functions;

(4) in accordance with the civil service laws, appoint and fix the compensation of such personnel as may be necessary to carry out the provisions of this section;

(5) obtain the services of experts and consultants in accordance with the provisions of section 3109 of title 5, United States Code;

(6) delegate an appropriate entity or independent agency, such as the National Science Foundation (NSF), to assist the Division with carrying out its obligations to appraise Qualified Platforms, Qualified Data and Information, Qualified Researchers, and Qualified Research Projects;

(7) accept and utilize the services of voluntary and non-compensated personnel and reimburse them for travel expenses, including per diem, as authorized by section 5703 of title 5, United States Code;

(8) enter into contracts, grants or other arrangements, or modifications thereof to carry out the provisions of this section, and such contracts or modifications thereof may be entered into without performance or other bonds, and without regard to section 3709 of the Revised Statutes, as amended (41 U.S.C. 5), or any other provision of law relating to competitive bidding;

(9) make advance, progress, and other payments which the Director deems necessary under this title without regard to the provisions of section 3324 (a) and (b) of Title 31; and

(10) make other necessary expenditures.

(h) The Director shall submit to the Chair, to the President, and to the Congress an annual report of the operations of the Division under this Act, which shall include a detailed statement of all private and public funds received and expended by it, and such recommendations as the Director deems appropriate.

SEC. 8. Enforcement

(a) Qualified Researchers who intentionally violate information sharing and privacy standards described in (a) shall be subject to both civil and criminal enforcement, under applicable federal, state, and local laws.

(b) The Commission is hereby empowered and directed to enforce the provisions of this Act, and violations of this Act by a Qualified Platform shall be deemed an unfair trade practice within the meaning of 15 U.S.C. § 45(a)(4).

(c) Whenever the Commission shall have reason to believe that a Qualified Platform has been or is in violation of any provision of this Act, the Commission may commence a civil action in a district court of the United States for an injunction against the Qualified Platform that the Commission believes has violated this Act. Remedies in an injunctive action brought by the Commission are limited to an order enjoining, restraining, or preventing any act or practice that constitutes a violation of this Act and imposing a civil penalty of up to [$10,000] for each violation, which shall accrue to the United States and may be recovered in a civil action brought by the Attorney General of the United States.

(d) In the event any enforcement action is appealed, the prevailing party in the action may, in the discretion of the court, recover the costs of the action including reasonable investigative costs and attorneys' fees.