FINAL REPORT

DEMOCRACY AND INTERNET GOVERNANCE INITIATIVE

TOWARDS DIGITAL PLATFORMS AND PUBLIC PURPOSE

HARVARD Kennedy School
BELFER CENTER
for Science and International Affairs
Technology & Public Purpose Project

HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

2023

Letter from the Co-Chair

Nancy Gibbs
Director, Shorenstein Center on Media, Politics and Public Policy

My first job out of grad school was with the International editions of TIME as one of its fact-checkers. We were responsible for the accuracy of every name, date, fact and figure we published, a responsibility based on the premise that there was some shared understanding of what constituted "truth" and "proof" until new evidence emerged. Fast forward to 2013, when as Editor in Chief I found myself leading a global newsroom through a time of momentous change—economic, political, technological, and epistemic. Print media was in decline, along with institutional trust more broadly; social media was on the rise, dividing audiences into parallel worlds of "alternative facts" in pursuit of power and profit. Every legacy newsroom wrestled with the ways technology was pushing us to rethink how we communicate information, engage with our audience, protect our writers and staff, and stay in business.

My co-chair, the late Secretary Ash Carter, who we sadly lost in October 2022, spent his career working to make America safer and more secure. When he took the Oath of Office as the United States' 25th Secretary of Defense, he swore to support and defend the Constitution of the United States against all enemies, foreign or domestic. He shared stories with me of how emerging technologies required the Defense Department to be agile, encouraging him and his team to view the threat landscape differently. From countering foreign interference to monitoring radicalization through online mediums, how we could keep the world safe needed to acknowledge the transformational nature of the digital ecosystem.

Secretary Carter and I joined forces to launch Harvard Kennedy School's Democracy and Internet Governance Initiative in 2021 because **we both shared the perspective that digital platform governance is one of the great issues of our time.** Today, our foreign and domestic enemies seek to weaken our democracy through the erosion of truth, the amplification of lies, and the weakening of the body politic. Our adversaries use our digital platforms to carry out their information operations; all too often, our platforms cannot, or will not, stop them. Additionally, there are limited mechanisms for consumer protection online, leaving individuals to deal with harassment, infringement of privacy, and exploitation.

Letter from the Co-Chair

And we have all experienced this change. The technological tools born from the Internet are complex, and challenge incumbent institutions in all corners of society. Parents must balance the benefits of social inclusion for their kids with the dangers platforms pose to mental health; political leaders leverage social media to reach, and sometimes manipulate, their audiences; Americans struggle to understand what is truthful online, all while being sucked into the addictive and polarizing vortex of social media.

It is long past time we act - to protect individual rights and freedom; to protect our public goods and information ecosystem; and, ultimately, to protect democracy. This final report is a culmination of our research and findings over the last two years. We hope the content contributes to the rich dialogue surrounding digital platform governance and helps us move from conversation to action.

Sincerely,

Nancy Gibbs
Lombard Director of the Shorenstein Center and
Edward R. Murrow Professor of the Practice of
Press, Politics and Public Policy

Table of Contents

Executive Summary

In an increasingly digital and interconnected world, platforms have emerged as powerful intermediaries that shape our online experiences, social interactions, and access to information. While platforms have brought numerous benefits, there is now an overwhelming recognition of their potential negative effects on individuals, society, and democracy. From the spread of misinformation and privacy concerns to cyberbullying and algorithmic biases, these harms demand a comprehensive and nuanced understanding, as well as mitigation strategies.

This paper serves as a summary report for the **Democracy and Internet Governance Initiative**, a two-year joint initiative between Harvard Kennedy School's Belfer Center for Science and International Affairs and Shorenstein Center on Media, Politics, and Public Policy. It delves into the rationale and components of a new risk-centered approach to analyze and address the negative impacts of digital platforms. It also explores the key dimensions that should be considered when assessing platform risk, including mental and physical health, financial security, privacy, social and reputational wellbeing, professional security, sovereignty, and strength of public goods.

The paper then underscores the necessity and value of comprehensive disclosure schemes in order to better understand the cause and effect of digital platforms and related products within the scope of a risk framework. Through the establishment of standards setting bodies dedicated to addressing itemized risks, we argue that this is the infrastructure necessary for sustainable regulation and self-governance that is dynamic and public purpose-oriented.

Executive Summary

Throughout each section, we aim to underline the significance of stakeholder engagement in the governance process. Platforms themselves, civil society organizations, policymakers, researchers, and users all have important perspectives, experiences, and jurisdiction needed to effectively mitigate harms of digital platforms. Engaging all stakeholders in a collaborative manner will enhance the framework's relevance, legitimacy, and practicality.

Ultimately, the risk-based approach and subsequent recommendations aim to contribute to a more informed and proactive approach to platform governance. By recognizing and addressing the risks associated with platforms—and confronting the pervasive information asymmetry problem, where technology companies withhold important data and information about user and community impact—we can work towards fostering a healthier, more accountable, and inclusive digital ecosystem that respects user rights, promotes responsible platform behavior, and safeguards the democratic principles that underpin our societies.

INTRODUCTION

MySpace, which launched exactly two decades ago in 2003, was the first online social network to reach a million monthly active users.[1] The platform, a friends-driven social media network, achieved this milestone in 2004, barely a year after its founding. MySpace demonstrated to the world how we could expand social interaction through the world wide web, allowing for new ways to discover music and art, share ideas with a large group of friends (location agnostic), and express yourself through digital artifacts.

In 2005, News Corporation (now Fox Corporation), sensing the potential business upside of online social networks, purchased MySpace for $580 million USD.[2] At that time, MySpace had 16 million monthly users–22 million in total–and was continuing to expand; its value at the time of the deal was approximately $327 million USD.[3] News Corporation, led by chairman and CEO Rupert Murdoch, was interested in leveraging a social platform like MySpace as a distribution outlet for Fox studio content. During Murdoch's ownership, the company expanded into the UK market, struck large deals with Google for advertising revenue, and attempted to build on its initial success while expanding Fox's viewership.

MySpace, despite Murdoch's investments, turned out to be a failure for News Corporation—and a failed venture overall.[4] The company's story, however, marks the beginning of an increasingly complex, and arguably convoluted, relationship between information and social networks. From Cambridge Analytica to the Arab Spring, social media has proven to be a useful tool for those interested in influencing real world events and the ways in which they are perceived. The MySpace story, though brief, shows the foreseeable risk born from the intermingling of digital platforms, information, and power. It also foreshadows something bigger: the potential of digital platforms to drastically change how we as humans interact with ourselves and the world.

Today, in 2023, MySpace is more or less out of the picture; Facebook is the largest platform in the world with over 2.98 billion monthly active users.[5] A number of other competitors entered the market over the last twenty years: TikTok, YouTube (acquired by Alphabet), Twitter, Twitch, Snapchat, WhatsApp (acquired by Facebook, now known as Meta), Instagram (acquired by Meta), and BeReal, to name a few.

---

[1] Ortiz-Ospina, E., & Roser, M. (2023, May 25). *The rise of Social Media*. Our World in Data. https://ourworldindata.org/rise-of-social-media

[2] U.S. Securities and Exchange Commission. (2005, July 18). News Corporation to Acquire Intermix Media, Inc. https://www.sec.gov/Archives/edgar/data/1308161/000118143105040705/rrd86058_6819.htm

[3] Adegoke, Y. (2011, April 8). *How myspace went from the future to a failure*. NBCNews.com. https://www.nbcnews.com/id/wbna42475503

[4] Ibid.

[5] Dixon, S. (2023, May 9). *Facebook mau worldwide 2023*. Statista. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide

The sheer power and wealth accumulated by these platforms is astonishing. Meta's market cap today is around $734 billion USD[6]; Alphabet, Google and YouTube's parent company, has a market cap of about $1.56 trillion USD.[7] That is larger than the GDP of many nations.

As these platforms gained momentum and integrated in the daily lives of billions of people all over the world, it became clear the information ecosystem was undergoing drastic change. Now, we see that large outlets no longer control a majority of content created online; the business model of platform companies requires news outlets to reconsider how they can reach and engage with their audiences; the rise of social media has led to a decline in local news, and has pushed for consolidation of larger outlets because of diminishing financial prospects; less reputable content creators are able to capture the attention of millions of users, allowing for the increased spread of misleading and false information at unprecedented speeds.

But the change goes well beyond the information ecosystem. Platforms undoubtedly have a significant impact on virtually every aspect of our lives. The average user spends nearly two and a half hours per day on various social media applications.[8] In 2021, during COVID, the average user spent over 1300 hours on apps like Facebook, Twitter, Instagram, and TikTok.[9] Due to the addictive features of social media and related psychological effects, the rates of teen depression have significantly increased.[10] Elder populations are subject to pervasive financial scams.[11] And because of limited capabilities for attribution, foreign actors have tampered and interfered with democratic processes, most notably the 2016 U.S. presidential election.[12]

As awareness of the negative consequences of social media grows, so does the demand for digital platform governance. The U.S. has been unable to make progress on specific legislation or governance schemes that could directly address the issue caused by digital platforms, though. Other countries and regions like Canada, India, and

---

[6] Yahoo! (2023, June). *Meta Platforms, inc. (Meta) stock price, news, Quote & History*. Yahoo! Finance. https://finance.yahoo.com/quote/META

[7] Yahoo! (2023, June). *Alphabet Inc. (GOOG) stock price, news, Quote & History*. Yahoo! Finance. https://finance.yahoo.com/quote/GOOG/

[8] Kemp, S. (2023, January 26). *Digital 2023 deep-dive: How much time do we spend on social media? - datareportal – global digital insights*. DataReportal. https://datareportal.com/reports/digital-2023-deep-dive-time-spent-on-social-media

[9] Suciu, P. (2022, November 9). *Americans spent on average more than 1,300 hours on social media last year*. Forbes. https://www.forbes.com/sites/petersuciu/2021/06/24/americans-spent-more-than-1300-hours-on-social-media/?sh=5170a6e42547

[10] Vidal C, Lhaksampa T, Miller L, Platt R. (2020, May). *Social media use and depression in adolescents: a scoping review. Int Rev Psychiatry.* 32(3):235-253. DOI: 10.1080/09540261.2020.1720623. Epub 2020 Feb 17. PMID: 32065542; PMCID: PMC7392374.

[11] Brancaccio, D. (2019, May 17). *Age of fraud: Are seniors more vulnerable to financial scams?*. Marketplace. https://www.marketplace.org/2019/05/16/brains-losses-aging-fraud-financial-scams-seniors/

[12] AP News. (2021, April 21). *Senate panel backs assessment that Russia interfered in 2016*. AP News. https://apnews.com/article/d094918c0421b872eac7dc4b16e613c7

Europe have made strides towards digital platform governance. The U.S., however, has been stuck in the political web of free speech, national competitiveness, and pro-market narratives - to the convenience of many companies.

The harms of platforms are only growing, though. Users globally are seeking change; it is time to act.

To that end, the Democracy and Internet Governance Initiative (DIGI) was launched in 2021 to research and deliver insights and recommendations to push the dialogue of social media governance towards action.

SCOPE & METHODOLOGY

Over the course of two years, DIGI focused on several key areas of public concern: diminishing press freedom, online extremism and radicalization, misinformation and disinformation, privacy, antitrust, mental and physical health, and financial security. We engaged over 100 subject matter experts and stakeholders through semi-structured interviews, written review periods, working groups, and briefing sessions.

Coupled with literature reviews and analysis supported by leaders in business, government, and civil society, we explored digital platform governance with one fundamental question at the center of our work: How can we, through a mix of self-governance and government regulation, create a dynamic and sustainable accountability system for digital platforms, namely social media, while preserving the benefits of platform technologies?

As a complementary lens for analysis, we applied history to understand the ways in which other innovation ecosystems are governed, with the goal of understanding and contextualizing the successes and failures of governance within other industries. This enhanced our research approach and better guided our recommendations for platform-specific schemes.

The goal of this report is not to present individual recommendations that may or may not address itemized risks, but rather to offer an updated approach to stakeholder conversations around problems caused by digital platforms. We also present process-oriented recommendations that could provide infrastructure to move governance from conversation and debate to targeted action.

Our research and methodology aimed to be non-partisan. The following content is a reflection and summary of our findings.

KEY INSIGHTS

These three key insights informed our final report:

1. Centering risk in our efforts to govern social media can provide us with an actionable north star.
2. There exists a deep information asymmetry problem – and there is only so much the government and civil society can do with limited information from technology companies.
3. We should not wait for the government to take action; as seen in other industries, we can leverage market dynamics to encourage private sector actors and experts across civil society to lead the charge on disclosures and the development of standards. This would simultaneously lay the foundation for informed and comprehensive government regulation.

Throughout the course of this paper, we expand upon these insights and make the case that a new risk-based approach to platform governance and a focused effort on platform disclosures can put us on the path to a dynamic multi-stakeholder governance scheme.

PART I: BACKGROUND

Over the course of two decades, digital platforms[13] have transformed the way we absorb information, perceive the world and ourselves, and engage with basic goods and services. They have undoubtedly created a number of positive outcomes. For example, social media has facilitated unprecedented global connectivity, enabling people to connect, share ideas, and build communities across geographical boundaries. It has also provided a platform for marginalized voices, activism, and social movements across the political spectrum, amplifying their reach and impact.

However, in the process of achieving scale and profitability, platforms have taken advantage of user data and infringed on privacy, encouraged overconsumption of digital content and goods, displaced incumbent media and entertainment outlets, and introduced a level of democratization to content production that has proven detrimental to modern democracies (ironic as it may be).

Luckily, the problem space related to digital platforms and the public is maturing. And as digital platforms continue to shift from being completely novel tools and products to being a pervasive and known part of society, observational studies by experts—sociologists, computer scientists, political scientists, economists, lawyers, and anthropologists alike—have more or less converged to say one thing: social media's negative externalities and companies' exploitive behavior is a known quantity issue and we need to act.

In this section, we provide an overview of the problems with digital platforms, as well as explore the solution space as it exists today, with a particular focus on the U.S. ecosystem, to serve as the foundation for Part II and Part III of the report.

PROBLEMS WITH DIGITAL PLATFORMS

One of the most well cited problems of social media and the online information ecosystem is the spread of mis- and disinformation.[14] Social media platforms amplify the reach of inaccurate or misleading content, leading to the spread of misinformation on a large scale. This can have severe consequences, including impacting elections through, for example, misinformation about voting stations to public health during crises. The NewsGuard's Misinformation Monitor in September 2022 found that for a sampling of TikTok searches on prominent news topics, almost 20 percent of the

[13] For the purposes of this report, we define *digital platforms* as "a software-based online ecosystem that facilitates interactions and transactions between users." For the scope of this paper, we primarily focus on social media; throughout the report, we use the terms *digital platform* and *social media* interchangeably.

[14] Misinformation is defined as "false information that is spread, regardless of whether there is intent to mislead." Disinformation is defined as "false information which is intended to mislead, especially propaganda issued by a government organization to a rival power or the media." Definitions based on Dictionary.com.

videos presented in the results contained misinformation.[15] For searches on topics ranging from the Russian invasion of Ukraine to school shootings and COVID vaccines, TikTok users are consistently fed false and misleading claims.

Digital platforms have been misused for more malicious purposes on the disinformation side. Bad actors, including state-sponsored entities and bots, exploit the algorithms and vulnerabilities of social media platforms to manipulate public opinion and sow discord. For example, the disinformation surrounding Russia's large-scale invasion of Ukraine in February 2022 marked an escalation in Russia's longstanding information operations against Ukraine and open democracies.[16] Propaganda and disinformation peddled by the Russian government have attacked Ukraine's right to exist. It has also accused Ukraine of being a neo-Nazi state, committing genocide against Russians, developing nuclear and biological weapons, and being guided by Satanism.[17]

Social media platforms have also provided an avenue for cyberbullying and online harassment, allowing individuals to target and harm others anonymously or under pseudonyms. This can have detrimental effects on mental health, self-esteem, and overall well-being, particularly among young people. A Pew Research Center survey of U.S. adults in 2021 found that 41 percent of Americans have personally experienced some form of online harassment in at least one of the six key ways that were measured. And while the overall prevalence of this type of abuse is the same as it was in 2017, there is evidence that online harassment has intensified since then.[18]

Digital platforms often collect vast amounts of personal data from users, which are exploited for targeted advertising, surveillance, or other purposes. Users are often not fully aware of the extent of data collection or have control over how their information is used, posing risks to privacy and individual autonomy. For example, on YouTube, content creators upload about 3.7 million videos in a single day.[19] The vast selection of videos on make-up tutorials, gaming, shopping hauls, product reviews, and educational videos are among the popular types of content that attract over 122 million users per day to the platform.[20] Viewership data, especially when linked with other information Google has on users, such as name, address, email, search data, map data, and more, provides a rich dataset for the company to use to target advertisements. It also has

[15] Brewster, J., Arvanitis, L., Pavilonis, V., & Wang, M. (2022, September 14). *Misinformation monitor: September 2022*. NewsGuard. https://www.newsguardtech.com/misinformation-monitor/september-2022/
[16] Bacio Terracino, J., & Matasick, C. (2022, November 3). Disinformation and Russia's war of aggression against Ukraine - OECD. https://www.oecd.org/ukraine-hub/policy-responses/disinformation-and-russia-s-war-of-aggression-against-ukraine-37186bde/
[17] Smart, C. (2022, July 2). *How the Russian media spread false claims about Ukrainian nazis*. The New York Times. https://www.nytimes.com/interactive/2022/07/02/world/europe/ukraine-nazis-russia-media.html
[18] Vogels, E. A. (n.d.). *The state of online harassment*. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/
[19] *YouTube stats: Everything you need to know in 2023!*. Wyzowl. (n.d.). https://www.wyzowl.com/youtube-stats/
[20] Ibid.

implications for the content that someone sees and experiences online, as the platform nudges people towards more engaging (or, rather, addictive) content.

The problem of financial scams on digital platforms is pervasive, with fraudsters taking advantage of the anonymity and wide reach of online spaces to target unsuspecting individuals. From investment schemes to phishing scams, these fraudulent activities can result in substantial financial losses and undermine trust in digital transactions.

Excessive use of social media and digital platforms can contribute to addictive behaviors and negatively impact mental health. The constant exposure to curated, idealized versions of others' lives, cyber comparisons, and fear of missing out can lead to feelings of inadequacy, anxiety, depression, and social isolation. Research has shown that young adults who use social media are "three times as likely to suffer from depression," putting a large portion of the population at risk for suicidal thoughts and behaviors.[21]

Digital platforms often use algorithms that personalize content based on user preferences and behaviors. While this can enhance user experience, it can also create filter bubbles and echo chambers, limiting exposure to diverse perspectives and reinforcing existing beliefs. This can hinder critical thinking, civil discourse, and understanding among different groups, exacerbating societal divisions.

Social media platforms often emphasize the culture of validation through likes, followers, and public recognition, leading to the constant pursuit of external validation. This can negatively impact self-esteem, body image, and overall well-being, as individuals compare themselves to others and strive for an unrealistic standard. In fact, according to an internal Facebook study, Instagram has harmful effects among a portion of its millions of young users, particularly teenage girls. Findings indicated that Instagram makes body image issues worse for one in three teenage girls. And among teenagers who reported suicidal thoughts, 6 percent in the U.S. traced them back to Instagram.[22]

Finally, radicalization and extremism campaigns are not new phenomena. What is new, however, is how the Internet can exacerbate the issue, providing speed and scale to the radicalization process.[23] For example, online social media platforms are playing an increasingly important role in the radicalization processes of U.S. extremists. In recent years, the number of individuals relying on these user-to-user platforms for the dissemination of extremist content and the facilitation of extremist relationships has

---

[21] *The impact of social media on Teens' Mental Health*. University of Utah Health | University of Utah Health. (2023, January 20). https://healthcare.utah.edu/healthfeed/2023/01/impact-of-social-media-teens-mental-health

[22] Wells, G., Horwitz, J., & Seetharaman, D. (2021, September 14). *Facebook knows Instagram is toxic for teen girls, company documents show*. The Wall Street Journal. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

[23] Jensen, M. (2018). Publication. *The Use of Social Media by United States Extremists*. National Consortium for the Study of Terrorism and Responses to Terrorism. University of Maryland.

grown exponentially. In fact, in 2016 alone, social media played a role in the radicalization processes of nearly 90% of the extremists in the PIRUS data.[24]

The intersection and advancement of problems

Further, the magnitude of harm grows at the intersection of some of these problems. For example, when privacy issues meet online harassment, it generates outcomes like doxxing.[25]

The problem landscape is also constantly evolving as technology itself changes. For example, when generative AI meets radicalization campaigns, it encourages things like deepfake videos that use uncanny depictions of threats to motivate and recruit new members to terrorist groups. The introduction of sophisticated generative AI into mainstream outlets and products adds a layer of complexity to social media harms by enabling the creation and dissemination of highly realistic synthetic content, making it harder to discern truth from fiction. It also creates challenges for content moderation efforts, as AI-generated content can evade detection algorithms, leading to the proliferation of harmful and misleading information. Additionally, the misuse of generative AI can facilitate the creation of targeted disinformation campaigns, further exacerbating the spread of online harms and undermining trust in digital platforms.

It is important to acknowledge how the evolution of digital technologies and services changes the manifestation of problems; privacy issues today in Web 2.0 may look different than privacy issues in Web 3.0, the Metaverse, or whatever our future digital environments look like. For that reason, it is important that the governance infrastructure we establish within the U.S. is dynamic and able to keep up with an ever changing risk landscape.

THE CURRENT SOLUTION SPACE

Section 230 reform has gathered a lot of attention as the primary solution to address the myriad of harms associated with digital platforms. However, Section 230, especially with its complicated political status, is not a silver bullet solution. And with the United States Supreme Court punting the debate back to Congress,[26] it is unlikely that we will see reform anytime soon.

Over the last several years, though, stakeholders and experts all across government, civil society, and industry have presented a wide range of robust and targeted recommendations to mitigate the harms of social media. As part of our methodology,

[24] Ibid.
[25] Doxxing, according to Oxford Languages, means "to search for and publish private or identifying information about (a particular individual) on the internet, typically with malicious intent."
[26] Liptak, A. (2023, May 18). *Supreme Court won't hold tech companies liable for user posts*. The New York Times. https://www.nytimes.com/2023/05/18/us/politics/supreme-court-google-twitter-230.html

we collected and analyzed more than 200 proposals related to U.S.-based platform governance across industry, government, and civil society.[27]

In order to highlight the priorities and public interest goals embodied within the existing proposals, we organized solutions thematically based on intent. The table below is a summary snapshot of the landscape. For a more extensive index of proposed solutions, we refer readers to the Democracy and Internet Governance Initiative's *Digital Platform Governance: Proposals Index*, a tool for researchers and other stakeholders interested in viewing the broader landscape of proposed solutions, primarily focused on U.S. intervention.[28]

Solution Bucket	Description	Examples
Improve transparency and accountability	Implementing regulations that promote transparency and accountability of social media platforms regarding their algorithms, content moderation practices, and data collection processes. This can involve requirements for clear disclosure of sponsored content, mechanisms for users to understand and control data sharing, and regular reporting on platform activities.	Platform Accountability and Transparency Act (PATA)[29]
Combat misinformation and disinformation	Encouraging platforms to develop and enforce policies against the spread of misinformation. This can include measures such as fact-checking, warning labels on disputed content, reducing the visibility of false information, and promoting credible sources of information.	Educating Against Misinformation and Disinformation Act[30]; "Label, Fact-check, and Delete False Content", as implemented by many platforms, including Twitter[31]
Counter extremism and radicalization	Proactive monitoring of vulnerable people and communities, collaboration with technology companies, community engagement, and educational initiatives aimed at promoting critical thinking and promoting alternative narratives that discourage extremist ideologies.	Interagency Working Group to Counter Online Radicalization to Violence[32]

[27] *Digital Platform Governance: Proposals index*. Belfer Center for Science and International Affairs. (2023, January). https://www.belfercenter.org/digital-platform-governance-proposals-index
[28] Access link: https://www.belfercenter.org/digital-platform-governance-proposals-index
[29] S.5339 - 117th Congress (2021-2022): Platform accountability and ... (n.d.). https://www.congress.gov/bill/117th-congress/senate-bill/5339
[30] H.R.1319 - American Rescue Plan Act of 2021. (n.d.-a). https://www.congress.gov/bill/117th-congress/house-bill/1319
[31] Twitter. (n.d.). *How we address misinformation on Twitter*. Twitter. https://help.twitter.com/en/resources/addressing-misleading-info
[32] Wiktorowicz, Q. (2013, February 5). *Working to counter online radicalization to violence in the United States*. National Archives and Records Administration. https://obamawhitehouse.archives.gov/blog/2013/02/05/working-counter-online-radicalization-violence-united-states

Protect data and privacy	Strengthening data protection laws to safeguard user privacy and limit the collection, storage, and sharing of personal data by social media platforms. Regulations can outline stricter consent requirements, data minimization principles, and enhanced user control over their information.	Online Privacy Act of 2021[33]
Address cyberbullying and online harassment	Enacting legislation that addresses cyberbullying and online harassment, providing legal recourse for victims and imposing penalties on perpetrators. This can include defining and criminalizing specific forms of online abuse, establishing reporting mechanisms, and fostering cooperation between platforms, law enforcement, and educational institutions.	"Platform Policies to Counter Online Harassment and Bullying", as Meta has implemented to protect public figures[34]; "News Organizations Training Journalists in Trauma Risk Management", as implemented by the BBC[35]
Strengthen digital literacy and education	Integrating digital literacy programs into educational curricula to help users develop critical thinking skills, media literacy, and responsible online behavior. Such programs can educate users about the risks of social media, teach them how to identify misinformation, and encourage ethical online practices.	To Establish the Digital Literacy and Equity Commission[36]
Increase platform responsibility and oversight	Establishing regulatory frameworks that hold social media platforms accountable for the content shared on their platforms. This may involve developing clear guidelines for content moderation, addressing hate speech, and setting up independent oversight bodies to monitor compliance and handle user complaints. Section 230 reform.	SAFE Tech Act of 2021[37]
Enhance interoperability, data portability, and user autonomy	Encouraging collaboration between social media platforms, researchers, and civil society organizations to share best practices, data, and technological solutions for addressing the risks associated with social media. This can involve industry-led initiatives, partnerships, and	Consumer Online Privacy Rights Act[38]

[33] H.R.6027 - 117th Congress (2021-2022): Online privacy act of ... (n.d.-b). https://www.congress.gov/bill/117th-congress/house-bill/6027/text

[34] *Advancing our policies on online bullying and harassment*. Meta. (2021, October 13). https://about.fb.com/news/2021/10/advancing-online-bullying-harassment-policies/

[35] BBC. (2021, March 11). *Psychological trauma support & trauma risk management (TRIM) - health & safety*. BBC News. https://www.bbc.co.uk/safety/health/trauma-support/

[36] H.R.6373 - 117th Congress (2021-2022): To establish the digital ... (n.d.-c). https://www.congress.gov/bill/117th-congress/house-bill/6373

[37] S.299 - 117th Congress (2021-2022): Safe tech act. (n.d.-d). https://www.congress.gov/bill/117th-congress/senate-bill/299

[38] S.3195 - 117th Congress (2021-2022): Consumer Online Privacy ... (n.d.-e). https://www.congress.gov/bill/117th-congress/senate-bill/3195

	information-sharing mechanisms to foster a collective response.	
Revitalize local and legacy news organizations	Revitalizing local news to provide reliable, context-rich, and community-focused reporting that fosters informed civic engagement and strengthens local democratic discourse. This can involve industry partners, civil society groups, and the government.	"Investments in community engagement and the revival of local journalism", as funded by the Knight Foundation[39]; Local Journalism Sustainability Act[40]

Through a thorough analysis of existing proposal and key areas of focus, we made a few observations:

1. The dearth of progress on federal legislation and federal action is not for a lack of trying. In fact, since the advent of social media, Congress has considered over 58 unique proposals, and at least 7 federal and local agency actions have been identified as potential solutions to digital harms.[41] However, Congress lacks the confidence and political will to move forward in what has become a deeply partisan debate.
2. Platform companies have tested the vast majority of proposals, meaning they are willing to try under the right circumstance - though none of the solutions have been fully implemented. Of the 70 initial proposals identified, just 6 proposals, or 9 percent, remain untested by industry, based on public information. It is worth noting that public information about company testing is limited and it is hard to understand whether solutions actually helped.
3. The solution space is (understandably) mainly reactionary, primarily focused on point-solutions for single issues. Generally, the conversation around digital platform governance has been fragmented. Advocates for reform are often subject matter experts focused on specific problem areas, such as mis- and disinformation or privacy. There has not yet been a comprehensive approach to addressing the issue space of digital platforms. Until that is achieved, the solution space will continue to be fragmented and reactionary.

The bottom line is that the solution space is rich with ideas. The fundamental issue however is that there is no clear way to understand which set of solutions will actually help address the current problems of social media, which is one of the reasons the landscape as it exists today does lean on more reactionary policy designs.

---

[39] *Investments in local news sustainability: Early learnings and insights*. Knight Foundation. (2022, May 21). https://knightfoundation.org/reports/investments-in-local-news-sustainability/

[40] H.R.3940 - 117th Congress (2021-2022): Local journalism ... (n.d.-b). https://www.congress.gov/bill/117th-congress/house-bill/3940

[41] Schultz, J. (2023, March 22). *Analyzing the landscape of solutions to social media's harms*. Belfer Center for Science and International Affairs. https://www.belfercenter.org/publication/analyzing-landscape-solutions-social-medias-harms

There are no current strategies to sandbox more multi-stakeholder solutions; only the ability to work with companies, for example, to build, test, and diagnose the effectiveness of new policies, products, and communication to mitigate harms or increase confidence in policy options through increased information on platform cause and effects. But the latter point operates under the assumption that companies are willing to disclose that kind of information. Unfortunately, they are not.

WHAT WE DON'T KNOW

Former Electronic Privacy Information Center (EPID) President Marc Rotenberg stated in 2018, "It's not clear why [Facebook], a company that has asked us to give up so much privacy, should be allowed to maintain so much secrecy."[1]

Throughout our research and literature review, we noticed a core issue: Although we have a lot of observational data aggregated by academics interested in understanding more about the profound effects of digital platforms, there is limited source data from companies about the real impact of products on users. Additionally, most user data generated and collected by non-company researchers often require self reported information, like time spent on an application or type of content viewed, leading to less than accurate information that then gets synthesized to understand addictive patterns, emotional responses, and content exposure. Companies collect and use this data to power their product engine but their blanket policy is that the information is proprietary. This makes it very hard for policymakers or researchers to collect insights about cause and effect on platforms, which is crucial for informed policymaking.

To reference a recent example, Meta banned news on Facebook and Instagram following the new Online News Act which recently passed in Canada. Facebook publicly stated that it is currently conducting several weeks of product tests to "end news availability in Canada" following the Canadian government's decision.[2] Meta said it would not release any results from product testing because it is proprietary information. (And although they are testing the impact of a news ban on user engagement, the company has taken a strict stance about fully banning news once the bill is active, with no room to negotiate with news outlets on content deals.)

Even PhD research fellowships funded by companies like Facebook, Twitter, Amazon, and Google to look into topics of concern, such as artificial intelligence and society or mental health and social media platforms, do not often get access to company data. Instead they are encouraged to pursue topics that companies ultimately know they are

[1] Romm, T. (2018, April 24). *Facebook's handpicked watchdogs gave it high marks for privacy even as the Tech Giant lost control of users' data*. The Washington Post. https://www.washingtonpost.com/news/the-switch/wp/2018/04/24/facebooks-hand-picked-watchdogs-gave-it-high-marks-for-privacy-even-as-the-tech-giant-lost-control-of-users-data/

[2] AP News. (2021a, February 25). *Facebook signs pay deals with 3 Australian News Publishers*. AP News. https://apnews.com/article/media-australia-9e4a02b5ddbd49c7eca310bdb71ddf80

ill equipped to tackle. Not unlike the tobacco industry, companies are incentivized to maintain the status quo. One of the best ways to do that is to keep information close.

Despite the deep information asymmetry that exists, there are several "known unknowns." These categories are essentially information deserts that need to be addressed:

"Long-term effects." Many studies have explored the immediate and short-term impacts of social media use, but the long-term effects on mental health, well-being, and social relationships are less understood. Longitudinal research is needed to examine the cumulative effects of prolonged exposure to social media and the potential long-term consequences for individuals across different stages of life.

"Individual differences and susceptibility." People's experiences and responses to social media vary widely. Understanding the factors that contribute to individuals' vulnerability or resilience to social media harms is complex. Factors such as age, personality traits, self-esteem, and pre-existing mental health conditions may interact with social media use in unique ways. Further research is necessary to better grasp these individual differences and how they influence the impact of social media on users.

"Causality and reverse causality." Establishing a clear causal relationship between social media use and specific harms can be challenging. While there is evidence of associations between certain social media behaviors and negative outcomes, it is often difficult to determine whether social media use directly causes harm or if pre-existing factors contribute to both social media use and negative outcomes. Additionally, reverse causality, where individuals with pre-existing difficulties may turn to social media more frequently, needs to be considered.

"Interplay of offline and online experiences." Social media is deeply intertwined with individuals' offline lives. However, understanding the complex interplay between online experiences and offline well-being is still an ongoing area of research. Further investigation is required to explore how social media interactions, both positive and negative, impact individuals' real-world relationships, social support, and overall life satisfaction.

"Differentiation of harms across platforms." While social media is often grouped together as a single entity, different platforms vary in terms of design, functionalities, and user demographics. Research needs to delve into the specific harms associated with different platforms to understand the unique risks and opportunities presented by different product types. Having an understanding of the differences in impact could actually lead us to understand how product characteristics can be healthier and less risky for the general public.

"Mitigation strategies and effectiveness." As efforts are made to address social media harms, understanding the effectiveness of various mitigation strategies is crucial. Research is needed to assess the impact of policy interventions, industry self-regulation, educational programs, and other initiatives aimed at reducing harm. Identifying the most effective approaches can guide the development of evidence-based interventions.

Research that is robust and rooted in source data is essential to deepen our understanding of the complexities surrounding social media harm and inform the development of targeted interventions and policies. At present, we do not have this basic ingredient, which arguably debilitates leaders from making decisions because without concrete insights on cause and effect, it is easy to fall into straw man arguments or weak claims—for example, that regulation would be detrimental to the U.S. innovation ecosystem, or that this is wholly a societal problem not a technological one—about why various solutions may or may not work.

In the following two sections, we elaborate on how we can develop the infrastructure for a cohesive domestic argument, focused on risk and improved process to address platform issues in an informed and multi-stakeholder manner.

PART II: ADAPTING THE U.S. NARRATIVE TO CENTER ON RISK

Drawing on the information presented in Part I, this section begins the outline how the United States can begin to unify conversations around platform harm and transform conversation and debate into action. First, we start with reorganizing the categories of harm to focus on public purpose dimensions. We then use the same public purpose dimensions to introduce an outcomes-based risk framework.

CATEGORIES OF CONCERN

The first area of concern is mental and physical health/safety. As referenced in previous sections, digital platforms have been associated with various risks to mental and physical health. Excessive use of social media can contribute to feelings of anxiety, depression, loneliness, and low self-esteem. The constant exposure to curated and idealized representations of others' lives can lead to negative social comparisons and a distorted self-perception. Moreover, digital platforms can contribute to sedentary behavior and a decrease in physical activity, which can have detrimental effects on physical health.

The second area of concern is financial security. Digital platforms can pose risks to individuals' financial well-being. Online scams, fraudulent activities, and identity theft are prevalent on digital platforms, potentially leading to financial loss and personal hardship. Moreover, platforms that facilitate peer-to-peer transactions or the gig economy may lack adequate safeguards for workers, exposing them to exploitative labor practices and precarious income.

The third area of concern is privacy. Digital platforms often collect and analyze vast amounts of user data, raising concerns about privacy. Users' personal information and online activities can be shared, sold, or exploited without their knowledge or explicit consent. This not only violates individual privacy but also enables targeted advertising, surveillance, and potential misuse of personal data by third parties.

The fourth area of concern is social and reputational risk. Digital platforms can pose risks to individuals' social relationships and reputations. Online harassment, cyberbullying, and the spread of false information can have profound negative effects on individuals' well-being and social interactions. Moreover, the viral nature of digital platforms can amplify the impact of reputational damage, leading to long-lasting consequences for individuals and communities.

The fifth area of concern is professional risk. Digital platforms can introduce professional risks and challenges. The increasing reliance on online job platforms can lead to precarious work arrangements, reduced job security, and inadequate labor protections. Furthermore, the public nature of online platforms can result in negative

impacts on individuals' professional reputations, with potential implications for career prospects and advancement. The blurring of personal and professional boundaries in online spaces can also create challenges in maintaining work-life balance and managing professional relationships.

The sixth area of concern is the risk to sovereignty. Digital platforms can pose risks to national sovereignty and democratic processes. The concentration of power in the hands of a few platform companies can influence public discourse, manipulate information, and impact political outcomes. Foreign interference, algorithmic biases, and filter bubbles can distort public opinion, undermine trust in democratic institutions, and erode the ability of societies to make informed decisions.

And last but certainly not least, the seventh area of concern is risk to public goods. Digital platforms can have both positive and negative effects on public goods. While they can facilitate access to information, knowledge sharing, and collective action, they also present challenges. The proliferation of misinformation, the erosion of traditional media outlets, and the dominance of algorithmic curation can hinder the availability of accurate and diverse information. Additionally, the extraction and concentration of economic value by digital platforms can undermine the sustainability of industries and sectors that contribute to public goods, such as journalism and creative content.

PROPOSED RISK FRAMEWORK

Below is a framework that outlines risk generated or exacerbated by the use of digital platforms. The risk framework focuses systematically on downstream outcomes for users i.e. it does not list misinformation and disinformation as a category of risk, but rather as a *driver* within examples of realized harm for the target of risk. For example, misinformation is a driver of risk to an individual's physical safety by informing them that "mugwort induces abortions." The outcome is harm to an individual's physical safety; pervasive misinformation on the platform can be one of the causes. In the following subsection, we elaborate on why reframing the conversation regarding platform risk could be beneficial in efforts to move conversation into action.

Note to Readers: The framework in its current state is not meant to be exhaustive; it serves as a preliminary proposal that can be adapted, expanded, and revised. The framework should be viewed as a dynamic tool that provides a systems level view on the state of risk for individuals and the greater public with regards to digital platform use.

Proposed Risk Framework

		Target of Risk	
		Individual Consumer	**Communal/Community**
Category of Risk	**Mental and Physical Health/Safety**	Individual exposure to danger, harm, or loss with regards to mental and physical health/safety. **Example:** Heightened depression and anxiety due to cyberbullying; poor personal health practices due to medical misinformation.	Community-wide exposure to danger, harm, or loss with regards to mental and physical health/safety. **Example:** National security and public safety risks, such as increased frequency of terrorist attacks by radicalized subgroups.
	Financial	Individual exposure to danger, harm, or loss with regards to financial matters. **Example:** Phishing scams through Facebook Marketplace that expose individuals to financial vulnerabilities.	Community-wide exposure to danger, harm, or loss with regards to financial matters. **Example:** Use of social media data for credit scoring, allowing for wide-scale precision marketing of predatory loans to vulnerable populations, as well as wide scale black boxed approaches for lending decisions. [1] [2]
	Privacy	Individual exposure to danger, harm, or loss with regards to privacy. **Example:** Data breaches or data vulnerabilities that are leveraged to doxx individuals.	Community-wide exposure to danger, harm, or loss with regards to privacy. **Example:** Long-term shifts in privacy norms, leading to less value placed on privacy overall for the public (allowing Chinese firms to collect intimate data on US users) which can be exploited during times of crisis.

1 The surprising ways that social media can be used for credit scoring. Knowledge at Wharton. (2014, November 5). https://knowledge.wharton.upenn.edu/article/using-social-media-for-credit-scoring/
2 Melancon, J. M. (2022, May 17). How social media posts could affect credit scores. UGA Today. https://news.uga.edu/how-social-media-posts-could-affect-credit-scores/

Proposed Risk Framework

		Target of Risk	
		Individual Consumer	**Communal/Community**
Category of Risk	**Social and Reputational**	Individual exposure to danger, harm, or loss with regards to social and reputational matters. **Example:** Reputational damage caused by pornographic material shared at scale (whether real or a deepfake).	Community-wide exposure to danger, harm, or loss with regards to social and reputational matters. **Example:** Reputational harm due to widespread misinformation, such as perception of the Asian community from misinformation about COVID origins.
	Professional	Individual exposure to danger, harm, or loss with regards to professional matters. **Example:** Social media reviews to screen and assess job candidates.	Community-wide exposure to danger, harm, or loss with regards to professional matters. **Example:** Algorithmic biases within professional platforms or hiring features, like LinkedIn's job board.
	Sovereignty	Individual exposure to danger, harm, or loss with regards to sovereignty. **Example:** Addictive product features, including hyper targeted content, that limit individual sovereignty over time and content exposure.	Community-wide exposure to danger, harm, or loss with regards to sovereignty. **Example:** Foreign interference in democratic processes, such as Russian interference in the 2016 U.S. presidential.

Proposed Risk Framework

		Target of Risk	
		Individual Consumer	**Communal/Community**
Category of Risk	**Public Goods**	Individual exposure to danger, harm, or loss with regards to public goods. **Example:** Through the development of communication norms, individuals are subject to decreased access to public goods, like important crisis communication, if they opt out of social media.	Community-level exposure to danger, harm, or loss with regards to public goods. **Examples:** Extinction of local news, and consolidation of other robust news outlets.

Within each area of risk, there are variables to contextualize risk, especially the level of risk that a given individual or subpopulation might be subject to. Variables for users could include the age, race and ethnicity of users, geographic region, and language. Variables related to the platform provider are also important for contextualizing risk. For example, the stage of the company and the business model are necessary to contextualize risk to understand what incentives the company has to act in the interest of their users or if they are willing to compromise user data, for example, in order to meet revenue goals.

For example, calling back to MySpace, when News Corporation sold MySpace to Viant, a digital advertising company, in 2011 for a mere $35 million. As part of the deal, MySpace shared all user profiles and data with Viant. Then in 2016, Time, Inc purchased Viant–and the old MySpace data along with it. This illustrates privacy risk, for example, when a company is looking to capitalize on remaining value during an acquisition.

Ultimately, risk needs to be assessed within context to ensure the right solutions are developed to address various forms and levels of harm to individuals and the broader community.

BENEFITS OF AN OUTCOMES-ORIENTED RISK-BASED APPROACH

A risk-based framework can help guide better solutions for digital platform governance by prioritizing resources and interventions based on the severity and likelihood of potential harms, and by centering the conversation around universal negative outcomes that are non-partisan. It provides us the opportunity to step out of the political arena and focus on the real consumer and community risks posed by digital platforms.

More specifically, a model that uses a comprehensive risk-based approach allows us to implement a systematic process to assess and identify the downstream risks associated with digital platforms. This involves considering a range of potential drivers of harms, such as mis- and disinformation and algorithmic bias, and translating those into public purpose oriented, downstream, probabilistic risk. As demonstrated through the risk framework, we argue that misinformation should be considered a *driver* of social and reputational risks, physical safety risks, and more; we should not center misinformation as a primary risk within the framework because it becomes more difficult to prove misinformation as an inherently harmful issue and it more explicitly shifts the argument to combating misinformation versus protecting free speech. By focusing on outcomes, and related drivers, we can conduct a root cause analysis and try, through technical and social means, to address the issue. Additionally, we benefit from a reframed debate around the physical health of children, for example, versus free speech. This latter form of the argument could yield better results as we try to move the needle on bi-partisan governance schemes.

And building on the point of better root cause analysis, like with many other risk models, this approach offers the opportunity to develop targeted mitigation strategies tailored to address the identified risks. This can involve a combination of regulatory measures, technological solutions, industry standards, and user empowerment initiatives - for example, implementing robust systems to discard high-risk content, promoting media literacy programs, and federally mandated disclosure requirements.

This can allow us to have proportionate interventions that focus on technological, social, and institutional best practices and standards that are enforced across the industry. Instead of applying one-size-fits-all approaches, such as broad Section 230 reform to address the myriad of risks associated with platforms, a risk-based framework allows for nuanced and context-specific solutions. This ensures that regulatory measures and enforcement actions appropriately match the severity and likelihood of harm.

The risk based model also allows for continuous monitoring and evaluation, and establishes mechanisms for ongoing monitoring and evaluation of risk mitigation measures. This involves assessing the effectiveness of implemented strategies, identifying emerging risks, and adapting governance approaches accordingly. Regular evaluation helps in refining policies and practices to stay responsive to evolving digital landscapes. It also encourages digital platforms to be accountable and transparent in their operations. This includes disclosing policies, practices, and algorithms related to content moderation, data handling, and user privacy. Increased transparency enables external scrutiny, promotes trust, and facilitates informed decision-making.

Finally, a risk-based approach can also encourage continued international cooperation by pulling the conversation out of the U.S. political context, and once again centering on the harms experienced by users all over the world. Digital platform governance is a global challenge and requires international cooperation. Encourage cross-border collaboration to develop common standards, share best practices, and coordinate regulatory efforts. The EU is already applying a risk model; the U.S. can collaborate on this effort and each can work to contextualize standards within their domestic jurisdictions. This can help address the transnational nature of digital platforms and ensure a consistent and coherent approach to governance.

Ultimately, this risk framework does not mean we can simultaneously mitigate all the risks at once. Instead, the framework gives us an ecosystem-wide perspective that allows us to appreciate the interrelation of issue spaces - and how we can balance tradeoffs between the myriad of harms (i.e. not just free speech versus content moderation, for example, which has a tendency to dominate the conversations).

PART III: DYNAMICALLY GOVERNING DIGITAL PLATFORMS

As stated in Part II, a comprehensive risk-based approach and framework allows us to implement a systematic process to assess and identify the downstream risks associated with digital platforms. However, in order to do that effectively, there are other variables that need to be considered, including disclosure, metrology, standards development, enforcement, and legal interpretation of risk and harm.

SOLVING THE PROBLEM OF INFORMATION ASYMMETRY (DISCLOSURES)

"Sunlight is said to be the best of disinfectants." - Louis Brandeis

As discussed in Part I, we do not yet fully know or have full information on the impacts of social media or other digital platforms with regards to consumer and communal welfare. There are no industry specific reporting requirements, and internal research is proprietary. Implementing disclosure requirements for digital platform companies is vital in promoting trust, safeguarding user rights, and ensuring a fair and informed online ecosystem.

One of the primary reasons for implementing disclosure requirements is, perhaps counterintuitively, to protect user privacy. Digital platforms often collect vast amounts of personal data from users, enabling targeted advertising and content customization. Disclosure requirements can ensure that platforms transparently communicate their data collection practices, providing users with clear information on how their personal data is collected, stored, and utilized. This empowers users to make informed decisions about their privacy and exercise control over their personal information.

Platforms can also foster trust by being transparent about their policies, terms of service, and community guidelines by implementing disclosure requirements. When users have confidence in the platform's practices, they are more likely to engage actively, share their thoughts, and contribute to the online community. Additionally, disclosure requirements have the ability to empower users to make informed decisions about their online experiences. By providing users with clearer information about the nature of the content they encounter, they can make more informed choices about the sources they trust, the information they consume, and the impact it may have on their well-being.

It is also important to remember that nearly every other industry in the U.S. has disclosure requirements. Some disclosure is voluntary by firms in order to build trust with users and gain a competitive advantage, and other disclosures are mandated by the federal government. These aid in government efforts to oversee various industries to ensure they are complying with the law.

The financial services industry, including banks, investment firms, and insurance companies, typically have federal disclosure requirements. These regulations aim to ensure transparency in financial transactions, protect consumers, and promote fair practices. The healthcare and pharmaceutical industries often have federal disclosure practices related to clinical trials, drug safety, adverse events reporting, and financial relationships between healthcare providers and pharmaceutical companies. These disclosures help ensure patient safety, ethical practices, and transparency in the healthcare sector. Industries involved in energy production, such as oil and gas, renewable energy, and utilities, have federal disclosure practices related to environmental impact assessments, emissions reporting, and compliance with environmental regulations. They are intended to promote sustainable practices and mitigate environmental risks. The food and agriculture industry also has federal disclosure practices regarding food safety, labeling requirements, and nutritional information. These disclosures aim to provide consumers with accurate and transparent information about the food they consume, ensuring their safety and facilitating informed choices.

Perhaps most prominent, the securities industry, regulated by the Securities and Exchange Commission (SEC), has federal disclosure practices to ensure transparency in financial markets. Publicly traded companies are required to disclose relevant financial information, business operations, risks, and executive compensation, among other aspects. Finally, the telecommunications industry often has federal disclosure requirements related to consumer rights, privacy policies, pricing, and service terms. They aim to protect consumers' interests and ensure transparency in the telecommunications sector.

In fact, digital platform companies have demonstrated in the past that they appreciate the value of disclosures to users in certain circumstances. In 2014, companies including Google, Facebook, Microsoft, and Apple pushed for the right to disclose information about National Security Letters and other requests they're required to comply with by law, including how much data they are required to share about user accounts.[46] The companies were levering disclosure to maintain trust with users amidst growing privacy concerns. Disclosures served as a useful tool to accomplish just that. To that end, during an era where individuals are losing faith in Big Tech platforms, disclosures offer upside not just for policymakers and regulators, but also for companies.

Ultimately, pushing for disclosure around consumer and public risk is vital for safeguarding user rights, promoting transparency, and fostering a fair and informed digital ecosystem. By ensuring privacy protection, transparency in algorithms, promoting fair competition, and facilitating user trust, disclosure requirements

[46] Hesseldahl, A. (2014, January 27). *Tech companies reach deal on disclosure of Security Data*. Vox. https://www.vox.com/2014/1/27/11622764/tech-companies-reach-deal-on-disclosure-of-security-data

contribute to a healthier online environment where users can make informed choices and engage meaningfully with digital platforms.

BUILDING A COMMON LANGUAGE FOR CAUSE AND EFFECT (METROLOGY)

Disclosure leads to another important variable: metrology. Metrology, the science and study of measurement, plays a crucial role in the development, understanding, and implementation of best practices and standards for any technology or innovation. Metrology is used in a variety of different industries including engineering, aerospace, manufacturing, energy, and healthcare. In engineering, metrology is used for structural analysis, and in manufacturing, it is used for quality control and to help cut down on wasted material.

Organizations like the National Institute of Standards and Technology's (NIST), housed in the Department of Commerce, are hubs and central coordinators for metrology work throughout the United States and the world. For NIST and organizations who participate in assessing risk and developing standards of new and existing technology, "advancing measurement science [enhances] economic security and [improves] quality of life."[47]

At a fundamental level, metrology ensures that measurements are accurate, precise, and comparable. Developing standards requires establishing reliable and consistent measurement methods to ensure uniformity and comparability across different industries, products, and services. For example, developing shared industry metrics on the impact of product changes to externalities related to social and reputational risk.

Something worth noting is that something called "social media measurement" does already exist.[48] However, its current application is primarily limited to evaluating the communication success of brands, companies, or other organizations on social media. The level of granularity in engagement measurement and cause and effect is impressive; if it were applied beyond marketing and with a public purpose lens, it could prove very powerful. Stakeholders could make data-driven and informed decisions to improve risk profiles across our public purpose dimensions.

DELEGATING ROLES FOR GOVERNANCE (STANDARDS, ENFORCEMENTS, ETC.)

Disclosure and metrology lay the foundation for real progress on platform governance. In this section we explore how, within a larger national infrastructure around digital

[47] *Metrology*. NIST. (n.d.). https://www.nist.gov/metrology
[48] Murdough, C. (2009) *Social Media Measurement*, Journal of Interactive Advertising, 10:1, 94-99, DOI: 10.1080/15252019.2009.10722165

platform governance, various stakeholders could lead and participate in the governance process.

Although this section is meant to provide a brief overview, we refer readers to a previous publication by the Democracy and Internet Governance Initiative, *To Break the Standstill of Social Media Governance, We Need Industry Standards*,[49] for an extended exploration of how standards and related government enforcement schemes prove opportunistic in the current market and political environments.

THE ROLE OF INDUSTRY AND CIVIL SOCIETY: DEVELOPING STANDARDS

Standard setting in technology industries refers to the process of establishing technical and operational specifications and guidelines for the design, development, deployment, and interoperability of technology products and services. This process involves bringing together industry stakeholders, such as developers, service providers, regulators, civil liberties groups, and standards organizations, to create and implement common technical and operational standards. Fundamentally, standards set out a common understanding among experts of "how things should be done if they are to be done effectively."[50]

Standards setting is that it is a known quantity process with a history of private sector engagement and success from the consumer welfare perspective. The ISO (International Organization for Standardization), which is an independent, non-governmental international organization, has a membership of 168 national standards bodies alone.[51] It works across a number of sectors including pharmaceuticals, energy technology, information security, and more. And although voluntary standards are non-binding, they often lead to mandatory standards enforced within a jurisdiction.

For digital platforms, the standards settings process offers a collaborative and ongoing medium to develop a common industry-wide language to measure and evaluate performance of online products and services, which is an important piece of the puzzle that is currently missing. It allows us to use a familiar and tested process to solve these somewhat novel problems, which has implications for global governance of digital platforms–not just domestic.

Industry-led standards development increases consumer confidence, builds trust with government, and can align with the fiduciary responsibilities of firms–all while supporting existing government and public interest initiatives. Though the process of standards development can be political, tedious, and imperfect, it serves as long term

[49] Access through link: https://www.belfercenter.org/publication/break-standstill-social-media-governance-we-need-industry-standards
[50] Hayns, S. (2020, August). *The importance of setting standards to support environment bill delivery*. Wildlife and Countryside Link. Retrieved from https://www.wcl.org.uk/the-importance-of-setting-standards.asp
[51] *About Us*. ISO. (2023, April 3). Retrieved from https://www.iso.org/about-us.html

infrastructure in which multi-stakeholder groups can collaborate on developing and implementing best practices.

And what is promising is that there have been scattered efforts to develop standards for social media companies. For example, the Global Alliance for Responsible Media (GARM) is a collaboration between advertisers, agencies, and social media platforms which aims to "develop and implement global standards for digital advertising, including issues related to brand safety, ad fraud, and hate speech."[52] The Sustainability Accounting Standards Board (SASB) had a Content Moderation on Internet Platforms initiative which was designed to "help companies manage the complex and evolving landscape of content moderation on internet platforms, while promoting user safety, privacy, and free expression."[53] Finally, the Digital Trust & Safety Partnership, which is part of the World Economic Forum's A Global Coalition for Digital Safety, has been developing best practices and assessments for digital service companies.[54]

Given the market dynamics with generative AI disrupting the standing of incumbent firms—as Microsoft aims to leap ahead of competitors like Google—and the impending regulatory shifts, the time is ripe for companies and civil society experts to collaborate on developing the infrastructure to develop, present, and implement best practices and standards. Using a shared framework of risk, experts can lay the foundation for robust government regulation that also aligns with market incentives.

THE ROLE OF GOVERNMENT: MANDATING DISCLOSURES AND ENFORCING STANDARDS

The history and founding of United States Pharmacopeia (USP), which is referenced in detail in *To Break the Standstill of Social Media Governance, We Need Industry Standards*, is just one example that demonstrates how self-regulation, like expert-led standard setting, can lead to smart government intervention that is based in technical and industry-based best practices. Standards can provide a framework for developing shared measurement, understanding of harms, and methods for protecting consumer and communal welfare, as was the case of the Pure Food and Drug Act.

Industry-led standards can also demonstrate to lawmakers and regulators that companies can self-coordinate to develop and enforce their own standards, which can allow them to avoid harsh or ill-informed government regulation. This also provides an incentive for industry practices to be transparent, accountable, and in line with social

---

[52] Advertisers, W. F. of. (n.d.). *Global Alliance for Responsible Media - About GARM*. WFA. Retrieved April 13, 2023, from https://wfanet.org/leadership/garm/about-garm

[53] *Content moderation on internet platforms*. SASB. (2022, July 6). Retrieved from https://www.sasb.org/standards/process/projects/content-moderation-on-internet-platforms-research-project/

[54] *Digital Trust and Safety Partnership*. Digital Trust & Safety Partnership. (n.d.). Retrieved April 13, 2023, from https://dtspartnership.org/

and environmental values in order to build the right level of trust with governments and consumers.

In the case of social media, the industry standards can lay the foundation for a few things: consumer safety and confidence, and legislative and regulatory input that has more depth and sustainability than lobbying. Ideally, once standards are created, documented, and applied, it becomes easier for Congress to codify those standards in law and appoint a regulatory office to enforce those standards, just like in the case of the FDA.[55] This benefits firms who opted into the standards pre-regulation because they are already in compliance. Additionally, domestic enforcement means that firms no longer have to worry about American competitors who opt to ignore the standards for the sake of, for example, first mover advantage.

The bottom line is that voluntary standard setting serves as a means to an end for government regulation. The long term play should be jurisdictional enforcement of standards via law, as well as an assessment by governments and civil society on whether the standards that are built are enough to effectively protect consumer and communal welfare. To do that effectively though, we need the foundation of some baseline shared interpretation of risk, disclosure, measurements, and expert-informed best practices to guide the dialogue. (Especially considering the limited scientific and technical capacity within the U.S. government for digital technology issues.[56])

THE ROLE OF THE JUDICIARY: UPDATING LEGAL INTERPRETATIONS

Last but certainly not least, as part of the broader governance infrastructure and ecosystem, updating the U.S. legal interpretation of threats to include non-physical harm in the context of social media governance is imperative. The digital age has expanded the scope of harm beyond physical injuries to encompass psychological, emotional, and reputational harm inflicted through online platforms. By recognizing non-physical harm, the legal system can effectively address the diverse range of harms propagated through social media.

Expanding the legal interpretation ensures the protection of vulnerable individuals, such as minors and marginalized communities, who are disproportionately affected by non-physical harm on social media. It acknowledges the unique challenges they face and helps establish appropriate safeguards and support mechanisms. Additionally, it acts as a deterrent, sending a clear message that harmful behavior online carries consequences and promoting responsible conduct in digital spaces.

---

[55] Former FCC Chairman Tom Wheeler, former Senior Counselor to Chairman at the FCC Phil Verveer, and former Chief Counsel of the US DOJ Antitrust Division Gene Kimmelman have a proposal to start a government agency to regulate digital platforms. Read more: https://shorensteincenter.org/new-digital-realities-tom-wheeler-phil-verveer-gene-kimmelman/

[56] Miesen, M., & Manley, L. (2020, November). *Building a 21st Century Congress: Improving STEM Policy Advice in the Emerging Technology Era.* Belfer Center for Science and International Affairs. Retrieved from: https://www.belfercenter.org/publication/why-us-congress-and-stem-experts-must-work-together

The importance of judicial interpretation is well demonstrated through a recent Colorado case: In June 2023, the U.S. Supreme Court overturned the stalking conviction, which implicated a man named Billy Counterman.[57] According to the case, Counterman sent more than a thousand Facebook messages to a singer-songwriter from Colorado. The musician claimed that the messages he sent made her fear for her safety. Causing sustained panic and fear, it upended her career as a performer. Based on a lower court decision, Counterman was sentenced to four and a half years in prison. However, when the state's Supreme Court took on his case, Counterman argued that the messages he sent did not meet the legal standard of a "true threat." To that end, he claimed the lower court decision was in violation of his First Amendment rights.

Ultimately, the Supreme Court voted 7-2 to send the case back to the Colorado courts, proposing a different standard of what speech is protected and what speech is a threat.[58] The Colorado court's upcoming decision will prove important in the interpretation of recklessness online and how we interpret a "true threat" online, which could have important implications for future cases. This is especially true since true threat historically has been rooted in proof of potential for real physical harm. When much of online harm is non-physical, we argue that the idea of a true threat necessarily needs to be recontextualized in the digital setting.

[57] Mohammad, L., Jarenwattananon, P., & Shapiro, A. (2023, June 27). *Supreme Court sets new standards for what constitutes "true threats."* NPR. https://www.npr.org/2023/06/27/1184655817/supreme-court-sets-new-standards-for-what-constitutes-true-threats

[58] Ibid.

CONCLUSION

Social networks and digital platforms have rapidly transformed our world: The sheer speed and scale by which individuals and institutions alike can share information has reshaped our information ecosystem; the new forms of service delivery including educational tools and financial services has made way for new forms of economic opportunity; the infrastructure that allows us to connect with people all over the globe has altered human socialization. And the outcomes of digital platforms are not all bad. However, as stated throughout this paper, the Democracy and Internet Governance Initiative joins a growing number of experts and institutions who are calling for earnest governance reform.

Over the past two decades, platform companies have gotten away with being under-regulated. They themselves have also chosen to largely opt out of conventional self-governance schemes like voluntary disclosures and standards setting when it comes to more consumer facing risk. Over the past two years, our Initiative was tasked to better understand the problem space, map the landscape of existing solutions, and develop a comprehensive and feasible plan to build domestic infrastructure for digital platform governance that acknowledged the necessity of a multi-stakeholder approach.

Throughout the literature reviews, working groups, expert interviews, and briefing sessions with policymakers and industry leaders, a few points became clear:

1. We have spent a lot of time debating point-solutions to a number of high priority problem areas, but we are operating almost blindly. Although disclosure–whether voluntary or required by law–may feel like a "timid"[59] intervention point, we need to remember that information is power. So far, technology companies have kept that power to themselves. And this is not arguing that the intuitive and observational findings of researchers are baseless. Rather, we are arguing that civil society groups, legislators, and regulators deserve to have access to information that is so core to our duties to protect individual rights and freedom, our public goods and information ecosystem, and, ultimately, democracy.
2. We need to focus on infrastructure and process if we want our governance scheme to be long lasting and dynamic, especially considering the rate of innovation in the digital sector. The digital landscape is constantly changing. With the introduction of sophisticated generative AI, for example, the threat landscape became even more complex thanks to even more convincing deepfakes and other synthetic media. Our governance mechanisms cannot be one off solutions, especially given the rate at which legislation moves. We need to focus our efforts on infrastructure and processes that are dynamic and agile.

[59] MacCarthy, M. (2022, November 1). *Transparency is essential for effective social media regulation*. Brookings. https://www.brookings.edu/articles/transparency-is-essential-for-effective-social-media-regulation/

3. We do not need to reinvent the wheel. There is nothing special about the digital platform industry that prevents us from applying well tested methods of governance and holding companies accountable.

Additionally, the conversation about digital platform governance has become politicized, somewhat sensationalized, and seemingly immovable. The purpose of this final report is to provide an updated framing of risk and a refreshed perspective on the infrastructure and processes grounded in feasibility analyses.

Throughout the paper, we lay out the case for sustainable infrastructure through disclosures, standards, and enforcement mechanisms working in unison, threaded together by a shared domestic perspective on risk. We aim to pull the conversation away from the mainstream political dialogue and towards something that can be implemented in a bi-partisan manner–and through the collaboration of business, government, and civil society.

Canada, the European Union, and the United Kingdom are taking the lead on industry regulation.[60] [61] It is time the U.S. acts, too. After all, it would be in the best interest of companies and the federal government to have the U.S. lead the charge on values-based governance schemes of American companies.

Ultimately, though, the primary motivation to act is that the cost of inaction is just too high. Without movement, our sovereignty, individual rights, public goods, and democracy are all at risk. As the late Secretary Ash Carter would say, we need to "land the plane." This report lays out the methods by which we can do just that; we call on leaders across business and government, in collaboration with civil society, to help move us towards real change.

[60] Chan, K. (2023, April 25). *Big Tech crackdown looms as EU, UK Ready New Rules*. AP News. https://apnews.com/article/tech-regulation-europe-tiktok-twitter-facebook-f9af8fdc69cab1e9a7ca836f5714bad7
[61] Coletta, A., & Vynck, G. D. (2023, June 22). *Meta says it will block news from Facebook, Instagram in Canada*. The Washington Post. https://www.washingtonpost.com/world/2023/06/22/facebook-meta-canada-bill-c18/

Acknowledgements

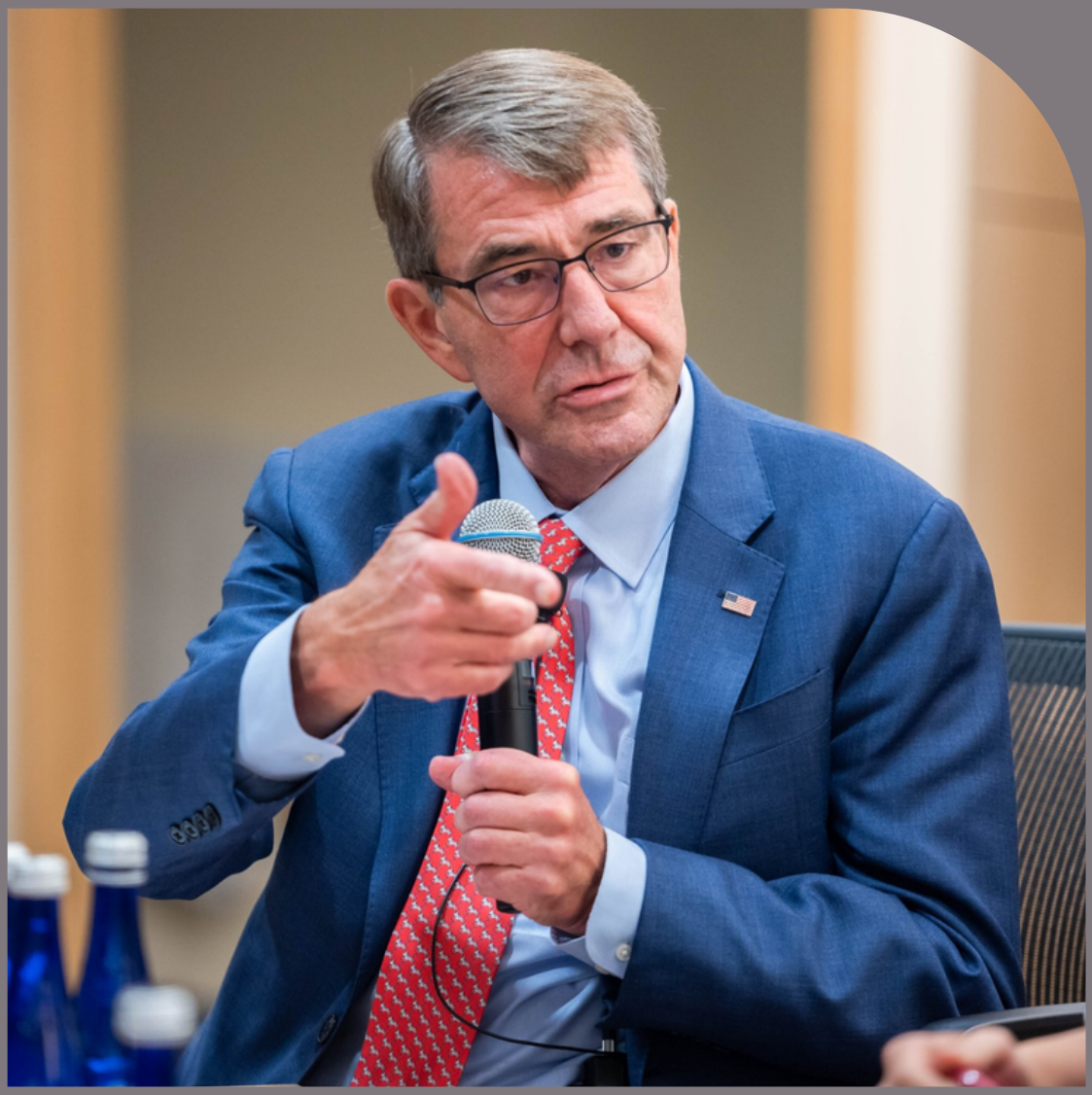
This report was made possible by contributions from the Harvard Kennedy School Belfer Center's Technology and Public Purpose Project and the Shorenstein Center's Technology and Social Change Project as part of the Democracy and Internet Governance Initiative (DIGI).

The Co-Chairs of this effort were Ash Carter (in memoriam), Former Director of the Belfer Center on Science and International Affairs and Nancy Gibbs, Director of the Shorenstein Center on Media, Politics and Public Policy. We would like to thank the numerous research assistants, project fellows, industry experts, civil society leaders, and others whose insights and feedback contributed tremendously to this final report.

HARVARD Kennedy School
BELFER CENTER
for Science and International Affairs
Technology & Public Purpose Project

HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

Ashton B. Carter



1954 - 2022

In memoriam

HARVARD Kennedy School

DIGI Final Report 2023