FALL 2020

Understanding Disinformation Solutions Landscape

Kelly Born

WILLIAM + FLORA
Hewlett
Foundation

TABLE OF CONTENTS

1. INTRODUCTION

The global information ecosystem has transformed dramatically in recent years. For almost a century, newspapers had served as the primary source of original reporting for all other media outlets. Between 2008 and 2019, employment at newspapers dropped by 51%.[1] By 2018, social media sites (e.g., Twitter, Facebook, etc.) had surpassed print newspapers in the U.S. as a more frequent source of news. And for the first time, in 2020, social media sites surpassed cable, network, and local TV as primary sources of political news for one in five Americans. And among U.S. adults under the age of 30, 48% get most of their political news from social media — suggesting that these numbers will grow considerably in the coming years as these younger populations reach maturity.

All things being equal, source would not matter. But concerns about the quality of information found on social media sites began well before Russian interventions in the 2016 elections. In 2015, a Pew study found that only "15% of those who get news from news organizations online find [the sources] very accurate." By 2018, over half of U.S. adults getting some news from social media said they expect the news they see to be largely inaccurate, and 74% of Americans believed that "the content people post on social media does not provide an accurate picture of how society feels about important issues." The salience of these concerns was clearly underscored when, first, Oxford Dictionaries named "post-truth" as the 2016 word of the year, followed in 2017 by Collins Dictionary's declaration of "fake news" as 2017's word of the year.

Yet while disinformation has spread widely, and trust in many U.S. institutions has declined — including trust in police, military, schools, business and religious leaders, and public officials — trust in science remains both the highest among these institutions, and has remained stable for decades. Recent surveys from Pew Research Center confirm that majorities still trust science. However, cracks in this foundation of trust are evident, including polarization of trust in science around specific issues including climate change, vaccines, GMOs, and, more recently, COVID-19. Recent polling during the pandemic is showing changes in this trust, and that trust may be increasingly correlated to political ideology and race for the first time.

In today's information environment, high-quality, socially and politically relevant, research-informed scientific information is often drowned out in a sea of noise and disinformation. Much work is needed to transform the new information ecosystem in ways that enable quality, evidence-informed information to thrive.

The following pages explore first the problem of disinformation, its origins, and the characteristics that make it distinctive from past propaganda challenges. This is followed by a discussion — and assessment — of the potential impact and feasibility of the many proposed solutions, including efforts to improve the quality of journalism, improve audience resilience to disinformation, or change social media company practices by improved content moderation, enhanced user controls, changes to privacy or liability law, antitrust enforcement, and more.

Finally, whereas both tech companies and government lack the political and economic incentives to address this problem, private foundations can play a pivotal role as neutral, outside agents of influence. The final sections thus provide an overview of key players in the field, followed by recommendations for what appear to be the highest-leverage opportunities for philanthropic support.

2. THE PROBLEM

A. Defining and Sizing the Problem

The term "fake news" was coined in 2016 to represent, technically, information that is completely and intentionally fabricated, and can be promulgated by anyone.[2] But the current "information problem" is multifaceted and reflects a range of concerns beyond blatantly "fake news," including:

DISINFORMATION
Intentionally false or inaccurate information that is spread deliberately, generally by unknown state officials, with a motive of undermining public confidence. (For simplicity, "disinformation" will be used throughout this report to generally describe all three forms.)

MISINFORMATION
Unintentionally promulgated, it refers to inaccurate information, which differs from propaganda in that it always refers to something that is not true, and differs from disinformation in that it is intention neutral.

PROPAGANDA
Information, generally promulgated by state officials, that may or may not be true, but that presents the opposing point of view in an unfavorable light in order to rally public support.

The disinformation problem is so difficult to address, in an information environment dominated by for-profit companies, in part because research has shown that people are psychologically predisposed to want to read news that is biased to reaffirm their preexisting beliefs and tribal identities. Research shows that mis/disinformation travels on average six times faster than the truth, as a direct function of how these types of socio-technical systems work: People gravitate toward content that is novel and emotionally engaging, with anger provoking the highest levels of engagement.

Meanwhile the business models of Facebook, Google, and Twitter are based almost entirely on engagement. More than 90% of their revenues are generated by selling advertisements.[3] These revenues are predicated almost entirely on the size of their user base, and the amount of time users spend on site interacting with content (i.e., engagement). Thus, because disinformation can drive engagement, problematic content can prove highly profitable for social media platforms. Scientific evidence seldom proves as novel or emotionally engaging, and thus struggles to gain traction on networks that are inherently social.

However, defining the problem of disinformation is complicated, for at least three reasons. First, scholarly definitions of the term "disinformation" differ widely, making it difficult to determine its prevalence.[4] Perhaps the most authoritative study on prevalence found that almost half of Americans (43%) visited known disinformation websites during the period immediately surrounding the 2016 elections.

In addition to conflicting views on how to define disinformation, and how to quantify its prevalence, concerns about the *exposure* of different groups to disinformation also vary. Some studies focus on whether specific articles were "seen and remembered."[5] Others are concerned with different communities' disproportionate exposure to disinformation, particularly across partisan lines.[6] And perhaps no group has been more exposed to political disinformation than African Americans. For example, race-related issues were found to be, by far, the preferred target of Russia's information warfare campaign in 2016: More than 66% of Russia's Facebook ads contained a term related to race, and 96% of Russian-funded YouTube content was targeted at racial issues and police brutality.

Others are concerned with measuring not just exposure, but *impact*. And here too definitions vary, with some studies attempting the methodologically challenging task of proving the causal link between the prevalence of scientific disinformation and its impact on people's subsequent behaviors.[7]

But others are concerned not just about content that might shift readers' views on any particular topic, but that trust in scientific evidence, global institutions, and the broader social fabric are being undermined by conflicting, racially charged, politically polarizing, inflammatory, or inaccurate content — even if the specific disinformation itself is not remembered. As Garry Kasporov eloquently put it, misinformation "annihilates truth."

B. Causes of the Information Ecosystem Decline

Social media platforms alone are not responsible for the deterioration in the U.S. (and global) information ecosystem. Several other important factors are at play:

GLOBALIZATION, ECONOMIC INEQUALITY

Since the 1970s global per capita GDP and economic equality have steadily improved.[8] Yet the distribution of these gains have been highly uneven — in the United States, the Gini coefficient (a measure of economic equality, with zero representing perfect equality and one representing perfect inequality) has grown from 0.37 in 1980 to 0.43 today. Economic inequality has often served as a significant catalyst for social unrest, social division, and a subsequent proliferation of narratives to explain the drivers of this inequality.

GLOBAL MIGRATION

Globalization is also correlated with increased migration and, according to Pew Research, "the sheer number of international migrants has never been higher."[9] Immigration is intrinsically tied to the search for jobs, often compounding economic tensions and sparking racial unrest.

BEHAVIORAL SCIENCE / HUMAN NATURE

Behavioral science confirms human tendencies toward motivated reasoning and confirmation bias — people are motivated to seek out and believe in information that reaffirms their worldviews and identities, particularly in polarized societies with significant political, economic, or social unrest.[10]

POLARIZATION

Driven, in part, by the above economic and racial tensions, the United States — and many other developed countries — have become sharply divided. Only a third of Americans now hold a mix of conservative and liberal positions, whereas in both 1994 and 2004, 49% did. Such divisions have been shown to decrease societal trust, which, in turn, contributes to the problem of disinformation.

CHANGES IN BROADCAST MEDIA LAW

At the same time that globalization took hold and economic, racial, and other divisions began to increase, the effects of the Federal Communications Commission's 1987 revocation of the 1959 Fairness Doctrine began to be felt. The Fairness Doctrine had required private media to provide what the agency regarded as balanced coverage of political events. Its revocation enabled the slow rise of highly partisan talk radio and cable TV, further polarizing America.

THE COLLAPSE OF NEWSROOMS AND RISE OF SOCIAL MEDIA

Finally, with the rise of social media, low-quality information has exploded into the information ecosystem at the same time that high-quality media sources have dramatically declined, as noted

above. While the decline of newspapers started well before the rise of the internet, the "creative destruction" wrought by digital technologies — whose primary source of revenue is the very same advertising funds that used to support professional journalism — has dramatically accelerated their decline.

Thus, while social media have unquestionably contributed to the current challenges in the information ecosystem, many other factors have helped to provide fertile ground for disinformation to take hold. And while addressing challenges with social media will not solve these underlying social problems, reducing the spread of disinformation can play an essential role in removing fuel from the fire.

C. How Digital Disinformation is Different

Disinformation and propaganda are not new — various forms have been around since the dawn of man. However, today's disinformation environment is markedly different in at least six ways. Identifying the interventions most likely to succeed in the modern information ecosystem requires and understanding of these core differences:

DEMOCRATIZATION

As Rand Waltzman, formerly of the Defense Advanced Research Projects Agency (DARPA), noted, any individual or group can now communicate with, and thereby influence, large numbers of others online. While there are numerous benefits, this also poses serious risks — beginning with the loss of journalistic standards of excellence. Without traditional institutional media gatekeepers, or any comparable "code of ethics" in the tech world, public discourse is no longer based on a common set of facts or ethical norms.

SOCIALIZATION

A direct byproduct of democratization: Rather than a system of "direct discovery," in which readers intentionally visit or receive news from the original source, we have moved to a system of "distributed discovery," in which audiences receive news from search engines, social media, and other platform products. These platforms and peer networks elevate content based on factors like clicks or engagement among friends, rather than accuracy or importance. Moreover, information that is filtered through networks of friends can result in an echo chamber of news that reinforces one's own biases (though there is considerable uncertainty about how serious a problem this represents). It also means that people who might otherwise consume more scientifically informed news are being inundated with more "popular" content, including extreme positions and falsehoods, which heighten the risk of misinforming wider swaths of the public.

ATOMIZATION

Individual news stories have now been divorced from brand or source. Previously, readers could easily distinguish between non-credible sources (e.g., the colorful and sensational tabloids in the supermarket checkout line, versus credible, longstanding local or national newspapers). Now, by contrast, an article shared by a friend or family member from the New York Times may not look all that different than one from a conspiracy theorist's blog. And, as the American Press Institute found, the original source of an article often matters less to readers than who in their network shares the link..

ANONYMITY *(in both information creation and distribution)*
Online news often lacks not only a brand, but also a byline. This obscures potential conflicts Online news often lacks not only a brand, but also a byline. This obscures potential conflicts of interest, creates plausible deniability for state actors intervening in foreign information environments, and creates fertile ground for bots to thrive. One 2015 study found that bots (software programs that can execute commands, reply to messages, or perform routine tasks such as online searches, either automatically or with minimal human intervention) generate around 50% of all web traffic, with as many as 50 million Twitter users and 137 million Facebook users exhibiting non-human behaviors. Of course there are "good" bots, say, providing customer service or real-time weather updates. But there are also plenty of bad actors "gaming" online information systems to promote extreme views and inaccurate information, lending them the appearance of mainstream popularity and acceptance. By 2018 estimates suggested that bots were behind 38% of all internet traffic.

PERSONALIZATION
Unlike their print, radio, or even television counterparts, internet content creators can A/B test and adapt micro-targeted messages in real time. "By leveraging automated emotional manipulation alongside swarms of bots, Facebook dark posts, A/B testing, and fake news networks," groups like the now infamous Cambridge Analytica can create personalized, adaptive, and ultimately addictive propaganda. In 2016, Donald Trump's campaign was measuring responses to 40-50,000 variants of ads every day, then tailoring and targeting their messaging accordingly.

SOVEREIGNTY
Unlike television, print, and radio, which are regulated by a patchwork of federal agencies, social-media platforms like Facebook and Twitter are today, for the most part, self-regulating — and are not very good at it (see below). While this is shifting, governments have been slow to identify and enact the policies necessary to address these problems, for reasons to be discussed.

Appreciating these differences is essential in considering the feasibility and impact of any interventions.

3. POTENTIAL INTERVENTIONS

No silver bullet exists, but a combination of the following interventions offer hope for dramatically improving our information ecosystem. While some appear fundamentally more promising than others, a suite of efforts will undoubtedly be needed to make progress.

In 2016, when the problem of disinformation exploded into the public consciousness, platforms, nonprofits, philanthropists, and government supported a wide range of interventions, broadly grouped by where in the information ecosystem they are focused: upstream, supporting the production of high-quality information (including scientifically informed research and journalism); midstream, on how information is distributed; or downstream, to support better audience consumption of information.

At that time, the majority of funds were directed upstream, to improve the quality of journalism produced, or downstream, supporting fact-checking, media/internet literacy, and other interventions aimed at information consumers. Less efforts focused on addressing the significant changes in information distribution or on increasing platform accountability — efforts which generally offer greater leverage.

The next section looks first at efforts focused upstream and downstream, exploring significant challenges with each, and then details midstream opportunities, working to directly improve the platforms.

A. Upstream Efforts to Improve Quality Information

U.S. philanthropists invest as much as $500 million annually to improve the supply of high-quality information, including funding journalism, attempts to improve newsroom revenue models, journalist education and training, media access and policy, audience engagement strategies, newsroom diversity, and more.[11] Yet despite the availability of scientifically informed journalism on issues ranging from the environment to vaccines, such information is often drowned out by noise and strategic dis/misinformation and propaganda.

In general, the existence of evidence-informed, high-quality information is a necessary but not sufficient condition, as misinformation often proves more emotionally persuasive than science. Addressing the overwhelming volume of low-quality content may be necessary before higher-quality content can rise to the fore.

While efforts to improve the quality of journalism/information writ large are unlikely to solve the online disinformation problem, two important journalistic efforts warrant mention:

First, **efforts to improve journalists' resilience to misinformation**: Journalists are susceptible to disinformation for two reasons: (a) the required speed of news cycles often allows limited time to authenticate information, and (b) because journalists remain significant gatekeepers to large audiences, they are most often the target of coordinated disinformation campaigns. Efforts to improve their resilience include training, tools, and other resources to help journalists adapt to the ever-changing threat of false information.[12] These efforts are essential because mainstream media outlets are still the largest potential amplifier of disinformation, either unwittingly or — as is particularly the case for disinformation initiated by political or other noteworthy figures — covering it as news.

Second, **efforts by investigative journalists to cover the problem of disinformation itself** as a beat: This includes extensive coverage by traditional journalism outlets like the New York Times and Washington Post, smaller for-profit outlets like Buzzfeed, and exceptional coverage by nonprofit groups, particularly ProPublica and the newly launched Markup, a digital-native data journalism outlet exploring the societal impact of big tech and algorithms. Support for tech watchdog journalism will prove essential to any future reform efforts.

B. Downstream (Audience-Facing) Efforts to Improve Information Consumption

There are now at least a half dozen efforts intended to help the public either correct or avoid dis/misinformation or propaganda, including:

1. "UNPLUGGING" OR "DIGITAL DETOXIFICATION"

There are many efforts aimed at getting users to refrain from or reduce their social media usage, with the #DeleteFacebook campaign likely the most popular. However, these appear unrealistic in the modern age, suboptimal as users also lose access to the many benefits of modern technology, and potentially counterproductive, as boycotting by "good actors" leaves "bad actors" to dominant the space.[13]

2. FACT-CHECKING

Groups like Factcheck, Snopes, and PolitiFact remain most prominent in the U.S., and Full Fact in the U.K. Efforts like Hoaxy take a slightly different spin, serving as a "search engine" for fake news, illustrating how claims spread on Twitter, and also fact-checking them. Many newsrooms have also launched their own efforts and partnerships: The Washington Post has its own Fact Checker initiative (home of the "Pinocchios"). In 2017 the BBC launched its own fact-checking effort — Reality Check — working with Facebook. By some estimates, more than 188 independent fact-checking groups have emerged, in more than 60 countries, in recent years, many now coordinated by the International Fact-Checking Network (IFCN) at Poynter. The tech platforms themselves are now also partnering with many prominent fact-checking groups in an effort to outsource these responsibilities, including Facebook's many efforts to support (but also pressure) global fact-checkers. Finally, in 2020, the Google News Initiative announced $6.5 million in funding support to support fact-checking organizations around the globe. A list of almost 40 of the more established, U.S.-based fact-checking initiatives can be found here.[14]

3. CREDIBILITY SCORING

While fact-checking is generally, though not always, done at the level of individual articles or statements, credibility scoring efforts often focus more on individual news outlets, and are intended to be used as credibility screens either for users, by tech platforms, or by advertisers seeking to avoid placing ads on "fake news" sites. Initiatives include the Global Disinformation Index, a crowd-sourced credibility rating system that outputs scores based on the evaluations of a network of "citizen scientists"; MediaBias Ratings, a web-based platform that produces media bias ratings based on a website's level and direction of bias, and also identifies questionable and conspiracy sources; or KnowNews, a browser extension developed through the Media Monitoring Africa initiative, which classifies news sites based on their credibility. A list of more than a dozen known credibility scoring efforts can be found here.

4. BOT/SPAM DETECTION APPLICATIONS

Significant efforts has been focused on creating "bot detection" tools such as Botometer, Botcheck.me, and Bot Sentinel, which users can install on their computers to help detect automated accounts.

5. NEWS LITERACY

At least 20 groups now work in the news literacy space, including the News Literacy Project, First Draft News[15], and the global development and education organization IREX, which operates in 100 countries around the world and runs Learn to Discern, a media literacy program. In addition, in October 2019, Facebook announced an initial investment of $2 million to support media literacy projects, and in June 2020 announced its partnership with the Poynter Institute to launch

the MediaWise for Seniors program (an expansion of a similar program for teens and college students).[16]

6. INOCULATION OR PRE-BUNKING

As an alternative to debunking or fact-checking (which occur ex-post and have limited proven efficacy) recent efforts at "pre-bunking" or "inoculation" preemptively warn and expose people to lower doses of misinformation, and have been proven to help cultivate "mental antibodies" against fake news.

All audience-facing efforts face at least one of several problems:

SCALE

Most public-facing efforts rely upon platform users to take action — to take a training, research an article, or download a new app. Given that today we can only get half of Americans to, on average, vote in a presidential election, the feasibility of efforts that require a great deal of users is questionable. Moreover, looking at fact-checking, estimates suggest that debunking false claims can take professional fact-checkers up to 15 hours, while a creator of disinformation can create false content in minutes, yielding a significant mismatch between the scale of the problem and the scale of fact-checking as a solution.[17]

PREACHING TO THE CHOIR

In addition to the challenges of reaching millions of readers with audience-facing interventions, unfortunately most interventions that require users to engage in trainings (such as pre-bunking or news literacy), download new internet applications (bot detection), or conduct additional research (credibility scoring) often fail to reach the most misinformed audiences, who are generally not as predisposed to visit fact-checking sites or enroll in related courses directly.[18]

MOTIVATED REASONING AND CONFIRMATION BIAS

Even when corrections do reach larger audiences, people often don't believe the correction. Behavioral scientists are consistent in their view that feelings and ideological affiliations often precede evidence in people's decision-making processes (i.e., evidence is used primarily to support preexisting feelings, and any dis-confirming evidence is often rejected, resulting in "belief persistence" for many, despite exposure to fact-checks). Views on the "favorability" of fact-checking also differ substantially by political party affiliation, and are higher among Democrats, as does confidence in the accuracy and nonpartisanship of news literacy or credibility scoring efforts (the latter of which are often faced with the question "who scores the scorers?").

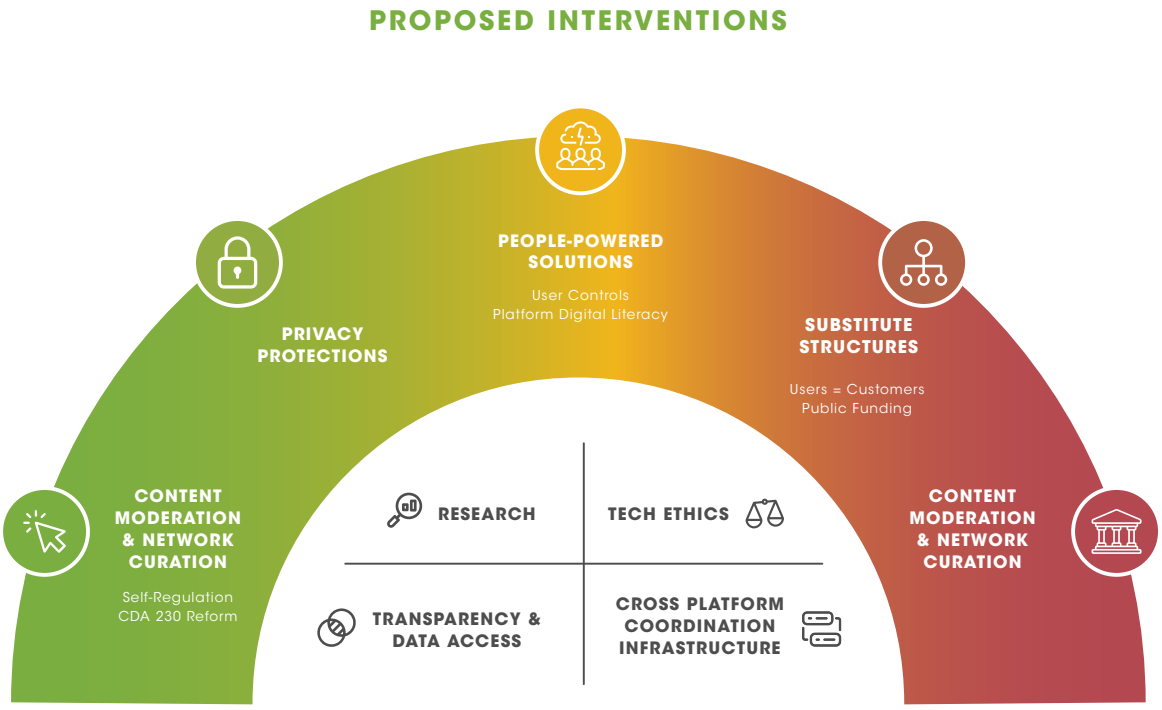
Of all audience-facing interventions discussed, pre-bunking and news literacy appear by far the most promising because they avoid the most common cognitive biases. Yet scale challenges remain daunting. Ideally, the scale of the interventions must match the scale of the problem — deploying pre-bunking and news literacy trainings through the platforms themselves offers the most feasible way to reach wider audiences quickly (as has been experiment with by David Rand at MIT). To enable this, much more research is required on the efficacy of different approaches, alongside advocacy to secure platform deployment of such interventions.[19]

C. Mid-Stream Efforts to Improve the Distribution of Information

Of all the changes in our modern information environment, it is the world's information distribution systems that have changed most radically with the advent of the internet and social media. Thus, it is here that there may be the greatest leverage and opportunity to affect positive change. A wide range of potential interventions to improve the role of social media companies have been proposed. The following chart illustrates:

- The key potential interventions proposed.
- The supporting infrastructure necessary to enable those interventions.
- The mechanisms through which these interventions might be achieved.

The following potential interventions begin with those most closely targeted to the problem of disinformation (content moderation), moving outwards toward interventions less closely related to the problem itself, and thus less likely to be successful and/or more likely to have unintended consequences (e.g., antitrust efforts).



1. IMPROVING CONTENT MODERATION & NETWORK CURATION

The most direct way to modify the social media ecosystem is through changing the platform policies that govern how users are allowed to engage, the algorithms that determine what content is elevated, or the kinds of groups or networks that are encouraged. Looking at *content moderation*, as mentioned, it is surprisingly difficult both to define and identify concerning content and, once identified, to determine how best to address it.[21] Equally problematic is the challenge of algorithmic *network curation*, where algorithmic recommendations can help to create a community of similarly minded users,[22] which can become fertile ground for future disinformation campaigns.[23] However, efforts of platforms to self-regulate by moderating problematic content, or reducing network formation, are unlikely to succeed so long as platform revenue models depend on high user engagement. And, in addition to economic disincentives, platforms have widely noted their philosophical reluctance to serve as "arbiters of truth," instead often espousing their commitment to free speech principles. And, even when platforms are willing to moderate content, recent data on the prevalence of COVID-related misinformation online suggests platforms often lack the technical or operational capacity to manage content effectively.

2. ENABLING USER CONTROLS

In an effort to avoid having to moderate content themselves, platforms continue to experiment with enhanced user controls, such as ad blockers, political ad filters, parental controls, URL filtering, and, most recently on Twitter, experiments with filters on who can respond to posts.[24] The European Commission's Digital Services Act, the most significant reform of European platform regulation in two decades, likewise seeks to provide "more user autonomy, choice and control."[25] Again, as with the above-noted concerns about any audience-facing interventions, the likelihood of these being widely adopted by users is low, given how often users simply accept platforms' default settings.[26]

3. INCREASING PLATFORM LIABILITY

Many have argued that, if platforms were held liable for the content they host, they would be heavily incentivized to remove problematic content. Calls for increased liability would necessarily require reforms to Section 230 of the Communications Decency Act of 1996[27], which effectively immunizes online service providers from legal liability for content posted by third parties.[28] While a key goal of CDA 230 was to promote innovation and protect free expression online, the provision has also been widely criticized for shielding technology companies from having to address harmful content, including disinformation, enabled by their services. Modifying Section 230 promises to increase platform accountability on the one hand, but runs the real risk of incentivizing technology companies to overcorrect and thus dramatically hinder free speech. Additional legal, technical, and theoretical research is needed to determine potential amendments capable of balancing liability with free speech.

4. ENHANCING PRIVACY AND DATA USAGE CONTROLS

Data about users' age, gender, demographics, political leanings, voting behavior, and other interests play a key role in the spread of disinformation by effectively targeting receptive audiences. By leveraging tools like Facebook's custom or look-alike audiences, motivated actors could, for example, identify users with similar characteristics to known climate deniers, and target them with related disinformation. By limiting access to the individual-level data that enables personalized ad targeting, data privacy laws "can render disinformation a weapon without a target." Without this detailed information that is used to algorithmically suggest content, disinformation has a higher chance of being lost in the noise. New privacy legislation, including the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act, though not directly intended to address mis/disinformation, enhance privacy rights and consumer protection by limiting how corporations can use personal data.[29]

5. PLATFORM-ADMINISTERED NEWS LITERACY

As discussed above, news literacy efforts hold considerable promise, but are complicated by the inability to scale, and likelihood of preaching to the choir. Deploying news literacy trainings directly through the platforms offers a solution better aligned with the scale of the problem.

6. NEW REVENUE MODELS

As discussed, the revenue model for social media and search platforms is predicated on selling advertisements, and thus user's attention.[30] In this sense, advertisers are the customer and users are the product.[31] Two alternative models are commonly discussed. First, a subscription model, with users, for example, paying for Facebook's services — obviating the need for Facebook to elevate the most engaging content.[32] However, this proves problematic globally, for less-developed countries that are heavily reliant on "free" platforms that often serve as the central venue for public discourse. And, given the demonstrated unwillingness of many readers to pay for newspaper subscriptions, this also seems unlikely to succeed. A second proposed approach is to support a "digital public square" wherein taxation of social media company revenues are used to support digital public infrastructure, akin to what has been done with public broadcasting in many countries.[33] However, such models face two complications: funding and usage. The United States has dramatically less appetite for funding of public broadcasting than many peer countries,[34] and that is unlikely to change given conservative resistance. Second, the assumption that "if we build it they will come" is unlikely to succeed in light of audience appetites for more confirmatory, polarizing outlets like MSNBC and Fox News. And, even in countries like the U.K., with substantial media budgets, problems of disinformation persist.

7. INCREASING COMPETITION THROUGH ANTITRUST, DATA PORTABILITY OR INTEROPERABILITY

Given the monopoly role that companies like Facebook, Google, and Twitter play in the information ecosystems, efforts to increase competition have arisen not just as potential economic remedies, but as remedies to platforms' growing political informational power. The argument in favor of antitrust suggests that if users had more alternatives to choose from, the threat to the broader information ecosystem posed by any single, small platform would be reduced.[35] However, while a temporary breakup might be possible, in the long term many argue that — because of so-called "network effects" — this may simply result in a competitor social media platform again gaining a monopoly position. As Stanford's Francis Fukuyama notes, "There is a very good chance that a baby Facebook created by such a breakup would quickly grow to replace the parent." [36] Moreover, breaking up the platforms may, conversely, make addressing disinformation and other harms more difficult, by creating more social platforms to monitor and regulate. Finally, while having many competing platforms might solve the large platform problem of exposing huge audiences to extreme, unscientific, or polarizing content, having many smaller platforms would create the problem of further siloing users with fringe beliefs into deeper echo chambers.[37] Much more research is required here to explore alternative antitrust structures better suited to address the platforms' political, rather than economic, harms.

D. Supporting Infrastructure

To enable many of the more than half-dozen interventions discussed above, a handful of supporting interventions are essential in helping to move the field forward, including tech ethics trainings, transparency and data access, longitudinal research, and coordinating infrastructure.

1. TECH ETHICS TRAINING

As the downsides of new technologies have become increasingly apparent, calls to begin training the next generation of technologists in ethical design principles have emerged.[38] Universities like Stanford, Harvard, and MIT — which helped produce some of Silicon Valley's leading technologists — have recently developed new courses, like the Ethics of Technological Disruption, as well as new centers, such as Stanford's Ethics, Society and Technology Hub.[39] These courses will prove critical.[40] That said, the time lag to impact of trainings like this is long, and in the absence of stricter legal restraints on the behaviors of technology companies, strong incentives remain for tech companies to pursue more profitable, less ethical paths. Additional work is needed to research which types of curriculum are most effective, and how to reach scale.

2. LONGITUDINAL RESEARCH, DEFINITIONS AND FRAMEWORKS

As leading scholars in the disinformation space note in the 2020 book "Social Media and Democracy," "Responding to an environment of panic surrounding social media's effect on democracy, regulators and other political actors are rushing to fill the policy void with proposals based on anecdote and folk wisdom emerging from whatever is the most recent scandal. The need for real time production of rigorous, policy-relevant scientific research on the effects of new technology on political communication has never been more urgent."

Much has been learned about the prevalence of disinformation and the mechanisms for its distribution, but much remains to be understood, including:

- Research to better define the problem(s) we are grappling with.
- Research examining the actual impact disinformation is having on people's beliefs and worldviews, including the impact on:
    - Trust in science generally, as well as on specific areas of intense public debate, including climate change, vaccines, COVID, and more.
    - Political divisions, in the U.S. and globally.
    - Racial and other key social divisions, including targeting of minority and underprivileged communities, which (at least in the U.S.) are among the most frequent targets of disinformation campaigns.
- Research examining the benefits, and potential unintended consequences, of the wide range of solutions discussed above.

3. TRANSPARENCY AND DATA ACCESS:

Today, research is severely hampered by a lack of access to data from social media companies. Platforms currently offer two primary avenues for insight: transparency reports and some (highly limited) direct data access.[41] While these limited reports and datasets provide some visibility into what users share on the platforms, what advertisers purchase, or what governments demand, they say almost nothing about how the platforms themselves are behaving as independent influencers of what content users are exposed to, or what groups users are nudged to join.

There are several potential mechanisms for securing data access:

- Academics could work with the platforms, through efforts like Facebook's Social Science One. To enable this, existing and proposed data privacy laws, like GDPR, would need to be clarified to ensure they allow data access for researchers, and platforms would have to be willing.[42]
- Governments could impose a "data tax" on the platforms — literally a tax paid through the contribution of some proportion of user-provided data into a repository for independent analyses, or in some other way mandate data access.
- Experts could work independently of the platforms to devise creative ways to collect social media data outside the platforms. This could include, for example, incentivizing users to install plug-ins that allow outside experts to monitor what content is being promoted in users' social media feeds in privacy-protected ways, as academics like Young Mie Kim at the University of Wisconsin, and journalism outlets including ProPublica through its Political Ad Collector, the New York Times, and the Markup (through their Citizen Browser partnership) have done.[43] Laws here would need to be changed to better protect developers of such efforts.

4. COORDINATING INFRASTRUCTURE

While significant funds have been invested in individual research institutes, there is surprisingly little coordinating infrastructure to support learning and collaboration across the field. Such infrastructure could be useful in (at least) three forms:[44]

- Across research entities, to support learning and reduce redundancies.
- Across platforms, to share information about changing disinformation actors, techniques, and narratives.[45]
- Between sectors, potentially including the platforms, academics, civil society, and/or the intelligence community.[46]

To support collaboration between tech companies and outside groups, several efforts have been attempted in the past, most notably the Global Internet Forum to Counter Terrorism (GIFCT)[47] and the Global Network Initiative.[48] However none are focused on addressing disinformation or the quality of the online information ecosystem.[49]

Such efforts face several hurdles. The attempts outlined above were entered into voluntarily by the platforms years ago, and there is some skepticism that voluntary collaboration could be secured in the current regulatory environment. The imbalances of power between tech companies and civil society actors in such settings have also proven difficult to overcome, as well as the lack of any enforcement mechanisms for any decisions made. More research is needed here to explore the pros and cons of differing potential structures.

IV. MECHANISMS FOR CHANGE

No silver bullet exists, but a combination of the above interventions and supporting infrastructure offer hope for dramatically improving the information environment. However, even if the most leveraged set of interventions can be agreed upon, much work will be required to see them implemented.

As with any field, implementation hinges on three main vectors for change: (a) voluntary change on behalf of the social media platforms themselves; (b) mandated changes by government, either through regulation or litigation; (c) public advocacy, generally aimed at motivating changes by the corporations or government:

A. CORPORATE SELF-REGULATION

Platforms have long pledged that they can themselves address disinformation, through improved natural language processing, machine learning, artificial intelligence, and more. This promise is appealing, in that it would enable quicker responses that are better informed by real-time changes in the tech world — versus government regulations, which may fail to keep up with the constantly shifting tech and external environments.

However, corporate self-regulation is fraught with both political and economic disincentives. First, intense societal divisions - be they religious, political, economic, or otherwise — make it *politically* fraught for platforms to serve in a role they frame as the "arbiters of truth." Second, even in circumstances where enhanced content moderation is politically feasible, the platforms have demonstrated, in the case of COVID-19, a lack of technical or operational capacity to successfully implement their desired policies or adequately self-regulate.[50] Finally, in cases less clear-cut than COVID, as discussed above, the underlying social media business model — with its priority on user growth and engagement — makes it is unlikely that platforms themselves will enact the fundamental changes necessary to address disinformation.

B. GOVERNMENT REGULATION AND LITIGATION

Government oversight is the optimal path, in that it is both democratically informed and enforceable. However, achieving optimal regulation in the current environment also appears challenging, for (at least) three reasons. First, and particularly in the United States, current levels of political polarization and legislative gridlock make it hard to get almost anything done, on this or any other issue of significant public concerns. Second, even if U.S. government were in a more functional state, government incentives to regulate tech are complicated by the fact that the Big Five tech companies — Facebook, Amazon, Apple, Microsoft, and Google — now represent more than 20% of U.S. GDP. Any proposed regulations that appear to threaten their profitability face notable challenges. And, at the same time, the tech giants have joined the ranks of the top 10 lobbyists in the United States.[51] And finally, even if government were able and willing, there are many values trade-offs inherent in tech regulation — between transparency and privacy, diversity of opinion versus a common knowledge base, free speech versus an accurately informed public — that complicate decision making.

Nevertheless, government regulation or litigation remain the most reliable, sustainable, enforceable mechanism for addressing these and other societal challenges democratically. And while U.S. Congress has been hamstrung in recent years, significant action is possible in the states — particularly California — and in Brussels. [52]

3. PUBLIC ADVOCACY

Finally, in the absence of public advocacy, neither governments nor tech companies are likely to pursue change, especially in light of the significant role tech platforms play in the U.S. economy. Boycotts have come from both users, including most notably efforts like #DeleteFacebook, and from customers (i.e., corporate ad buyers) through efforts like #StopHateForProfit and Sleeping Giants — though, of course, these advertiser boycotts were themselves inspired by their customers.[53] While boycotts are unlikely to prove sustainable, continued public pressure will prove essential in ensuring ultimate change. The extensive journalistic coverage of the disinformation challenge, noted above, has played a critical role in fueling this pressure.

However, it is worth noting that many of the most prominent nonprofit efforts to hold the tech platforms accountable are, themselves, funded (at least in part) by the platforms.[54] Such tech funding is both essential — given the lack of other funding sources — and of course problematic, given the conflicts of interest such funding presents. Again, this highlights the dire importance and great opportunity facing private foundations, which have the capacity to support both research and advocacy.

V. LANDSCAPE: THE SIZE AND SHAPE OF FIELD

Despite the significant attention paid to the problem of disinformation in recent years, there are relatively few active funders in the space. Led by a $10 million investment in disinformation research from the Hewlett Foundation starting in 2016, and followed by a $50 million investment, to almost a dozen U.S. universities by the Knight Foundation in 2019 to support research "to better understand how technology is transforming our democracy and the way we receive and engage with information," more than 50 funders have made grants in this space since 2016. However, the majority of these funders have made relevant grants totaling less than $100K. Only an estimated nine funders that have invested up to $1M in the disinformation field, including:[55]

GRANTMAKER NAME	AMOUNT INVESTED SINCE 2016
THE KNIGHT FOUNDATION	$ 42,787,624
THE WILLIAM AND FLORA HEWLETT FOUNDATION	$ 10,200,000
THE UNITED STATES DEPARTMENT OF STATE	$ 5,603,740
THE DEMOCRACY FUND	$ 3,569,398
THE FORD FOUNDATION	$ 3,160,000
UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT	$ 1,570,000
THE FOUNDATION TO PROMOTE OPEN SOCIETY	$ 1,528,954
KING BAUDOUIN FOUNDATION UNITED STATES	$ 1,497,861
OPEN SOCIETY INSTITUTE	$ 998,850

Of these top funders, three (the United States Department of State, the King Baudouin Foundation, and the Ford Foundation), invested heavily internationally, with cumulative non-U.S. grants of $5.3 million.

These grants have supported more than 100 organizations doing work related to fake news, misinformation, or disinformation (with some nonprofits receiving as little as $500 in total relevant grants). Only 24 organizations have received more than $500,000 in funding since 2014 (15 universities and nine think-tanks / other entities):

ACADEMIC INSTITUTION	AMOUNT
NEW YORK UNIVERSITY	$ 5,144,440
CARNEGIE MELLON UNIVERSITY	$ 5,000,000
GEORGE WASHINGTON UNIVERSITY	$ 5,000,000
THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL	$ 5,000,000
UNIVERSITY OF WASHINGTON FOUNDATION	$ 5,000,000
SOCIAL SCIENCE RESEARCH COUNCIL	$ 4,544,924
THE UNIVERSITY OF TEXAS AT AUSTIN	$ 3,434,958
PRESIDENT AND FELLOWS OF HARVARD COLLEGE	$ 3,325,000
TRUSTEES OF INDIANA UNIVERSITY	$ 3,000,000
BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY	$ 2,900,000
DUKE UNIVERSITY	$ 1,308,000
CITY UNIVERSITY OF NEW YORK	$ 1,000,000
UNIVERSITY OF WISCONSIN-MADISON	$ 1,000,000
UNIVERSITY OF WASHINGTON	$ 754,046
UNIVERSITY OF NORTH CAROLINA	$ 600,000

THINK TANK/OTHER NONPROFIT	AMOUNT
NATIONAL ENDOWMENT FOR DEMOCRACY	$ 2,716,050
GERMAN MARSHALL FUND OF THE UNITED STATES	$ 2,650,000
FIRSTDRAFT	$ 1,700,000
UNICEF	$ 1,570,000
CENTER FOR DATA AND TECHNOLOGY	$ 1,513,000
DATA AND SOCIETY RESEARCH INSTITUTE	$ 1,250,000
COMMON SENSE MEDIA	$ 800,000
MEDIA MATTERS FOR AMERICA	$ 799,500
OFICINA LEGAL DEL PUEBLO UNIDO, INC.	$ 500,000

While significant funds have gone to support research in this space, the vast majority of that research remains focused on tracking the prevalence of disinformation, with relatively less focused on exploring its impact, or the impacts (and potential unintended consequences) of the interventions discussed above. And much less funding still has gone to supporting the key infrastructure and advocacy investments also discussed here.

When it comes to improving the information ecosystem, it is clear that there will be no single, comprehensive solution. Instead many interventions — large and small, and addressing the problem from many angles discussed above — will be required. Thus there is room for many different types of funders to contribute to solutions, from angles that align with their own strategies and grantmaking histories.

VI. CONCLUSION

Misinformation, disinformation, and propaganda are not new. But social media and search platforms are. Many of their central characteristics — including speed, scale, and democratization of information — are features, not bugs. The elevation of more popular click-bait content is a lamentable byproduct of a design that prioritizes user growth and engagement above quality content.

As foundations consider how to address the problem of disinformation, interventions focused on improving the role of technology platforms appear most promising. Focusing on the tech platforms alone won't solve the information problem, any more than the platforms alone created it. But work in this area is essential if we are to deploy solutions aligned with the scale of the problem.

Of the many potential efforts outlined above, five offer the most leverage, and are organized roughly in order of their anticipated impact and feasibility:

A. TRANSPARENCY

Efforts to provide privacy-protected data to credible researchers will prove essential in enabling the analyses necessary for scholars, advocacy organizations, governments, and all other relevant actors to hold platforms accountable. Without these data, little about the platforms can be understood.

B. RESEARCH

Significant research is still needed to better define the problem(s) of disinformation; to understand the impact it is having on people's beliefs and worldviews; and to examine the benefits, and potential unintended consequences, of the wide range of solutions discussed above. Moreover, because the techniques of bad actors, the policies of governments, and the algorithms and practices of the tech platforms are constantly evolving, the need for this research is far greater than current funding can support, and is unlikely to diminish anytime soon.

C. NEWS LITERACY

Of all audience-facing efforts, pre-bunking and news literacy appear by far the most promising. Yet surprisingly little is known about the efficacy of different approaches, and scaling such efforts remains daunting. Significant research is needed to understand which approaches are most effective, followed by organized advocacy to ensure such interventions can be deployed at scale, ideally through the public school system and/or social media platforms.

D. TECH ETHICS

To train the next generation of tech leaders, as well as the current generation of leadership in tech, government, and civil society, additional work is needed to research which types of curriculum are most effective, and how to reach scale.

E. COORDINATING INFRASTRUCTURE

Finally, few mechanisms exist to support coordination within this new field of disinformation research, or between relevant parties in academia, civil society, government, and the private sector. While calls for such infrastructure have been made, no concrete proposals have yet been developed or analyzed. Research on this front could provide significant leverage.

For each of the above efforts, a promising first step would be to commission focused, collaborative research — in the form of an edited volume and/or workshop series, costing less than $100,000 — bringing together leading experts in each field to identify what the empirical literature has to say about each topic, what

recommendations can, at present, confidently be advocated for, and what key knowledge gaps remain. From there, public advocacy (to government, platforms, or academic institutions) would be needed to secure change.

The public square has moved online, societies have begun to fragment along racial, religious, and economic lines, and social media — rather than credentialed journalists — now hold significant power to not only (mis) inform the public and highlight key issues, but to unite like-minded strangers galvanized by new (and at times scientifically inaccurate) ideas and enable the formation of new social groups with their own distinct information ecosystems.

This power to inform, unite, and organize, at a global scale, is unlike any society has seen before, and offers many benefits. At the same time, the algorithms governing the new public square are designed to maximize user engagement and revenue growth. And users are known to engage most with novel and emotionally stimulating content — the kind of content perhaps least likely to be grounded in science. Yet such scientifically grounded information is essential to society's ability to think critically, openly, and creatively; to reason soundly; and to make informed collective decisions about the many critical issues facing our global future. Unless significant work is done to improve the current information environment, it is hard to imagine a future in which evidence-informed, scientifically grounded facts will play a central role in the public consciousness.

[1] Between 2008 and 2018, newsroom advertising revenue fell from $37.8 billion to $14.3 billion, a 62% drop.

[2] Buzzfeed's then media editor, Craig Silverman, is credited with coining the term "fake news" when he noticed a stream of entirely fabricated stories originating from a small town in Macedonia.

[3] In 2019, Facebook revenues hit $70.7 billion, up from $55.8 billion the previous fiscal year, 98.5% of which were generated from ad sales. Meanwhile Google's 2019 revenues were at $162 billion, while Twitter revenues are comparatively small at $2.5 billion.

[4] In considering definitions: To count as "disinformation," must the entirety of the article be falsified, or does an article with some misleading facts woven in alongside accurate content, or predominantly factual content taken out of context and portrayed in an inflammatory or polarizing light, count? Should scholars be measuring this content at the level of the individual article, the author, or the entire website (the latter two of which may offer a mix of accurate and misleading articles)?

[5] One prominent study found that the average American adult saw and remembered slightly more than one fake news article in 2016.

[6] A study found that 62% of visits to "fake news" cites surrounding the 2016 elections came from the 20% of Americans with the most conservative information diets.

[7] A notable case study from Pakistan in 2019 found that, immediately after false rumors about the safety of the polio vaccine began spreading on social media, vaccine refusals jumped from 256 instances in March to 88,000 in April.

[8] While economic globalization began with the advent of trade, as early as 6500 BCE, modern globalization was stymied by protectionist economic policies following World War I, and did not resume until the 1970s as technological advances exploded, and governments began to adopt more market-oriented reforms. The global Gini coefficient has fallen from 0.80 in 1980 to 0.65 today.

[9] By 2017, almost 50 million people living in the United States were born in other countries.

[10] Research has also shown that audiences that believe in any single conspiracy theory are in turn more likely to believe in other conspiracies and pseudoscience.

[11] Top funders include the Knight Foundation, Freedom Forum, the Ford Foundation, Democracy Fund, and the Gates and Hewlett foundations. Top grantees include Columbia University's Journalism School and the Annenberg School for Communication, Poynter Institute, the American Press Institute, and Solutions Journalism Network, which all do incredible work educating journalists, as well as nonprofit newsrooms like ProPublica and the Center for Public Integrity, which do remarkable reporting supported by philanthropic funds.

[12] Groups like First Draft News deliver training programs and collaborative reporting projects for journalists to help them navigate and halt the spread of misinformation and disinformation online; Harvard's Shorenstein Center on Media, Politics and Public Policy works to equip news executives with knowledge and frameworks to address media manipulation; and Stanford's Cyber Policy Center has developed tools to help journalists determine the source behind disinformation campaigns, as well as practices to cover hack and leak efforts and other manipulations of the mainstream media.

[13] These efforts are explored in more detail in Cal Newport's book "Digital Minimalism," and in Jaron Lanier's "Ten Arguments for Deleting Your Social Media Accounts Right Now."

[14] Other groups like First Draft News, founded in 2015 as a coalition of newsrooms and social media platforms, are working to improve practices, standards, and technology for debunking problematic information and verifying eyewitness media online.

[15] New to the space, First Draft recently deployed a text messaging-based digital literacy course.

[16] "Despite their technical prowess on smartphones and social media," recent studies have found that even digitally literate students struggle significantly to sort fact from fiction, and are easily fooled by misinformation even when given the time and resources to fact-check the material. (In one telling experiment, two-thirds of college students failed to identify that a misleading story was published on a satirical website and was not reliable. In a second experiment, more than nine in 10 students failed to realize that the website "purporting to provide unbiased information…had actually been established by a public relations firm funded by an interest group" opposing the issue at hand.

[17] Looking at news literacy, similar scaling challenges exist, with many approaches deployed to date relying on time-consuming in-person trainings (for which longitudinal efficacy data are also scarce).

[18] To overcome this, some scholars are experimenting with efforts to serve up fact-checks, pre-bunks, or news literacy prompts in "real-time," through social media platforms — though additional research is needed to test efficacy of these "smaller dose" interventions, and platforms would need to be willing to deploy them.

[19] RAND has done perhaps the most comprehensive assessment of media literacy interventions, finding that there is wide variation in both the definition of media literacy, the goals of the programs, and mechanisms for analyzing results.

[20] For this reason, platforms have avoided making judgments based on the *content* itself, instead preferring to focus on understanding the *actor* behind the content, or the *behaviors* or techniques those actors are employing to artificially amplify their content — creating fake accounts, running bot networks, or micro targeting in discriminatory ways. This framework is now referred to as the "ABCs of disinformation" — actors, behaviors, and content.

[21] Once problematic content — paid or unpaid — is identified, platforms must determine how to deal with it. Early calls asked the platforms to delete problematic content, again fueling free-speech concerns. Potential platform interventions are now more nuanced: In addition to deleting content, platforms can demote it, disclose the source behind it, delay content that has reached a certain threshold of virality until it has been verified, dilute it amid higher-quality content, provide disclaimers (for example by attaching fact-checking, source information, or additional context on the topic), deter profit-motivated actors from placing it (by changing ad placement or compensation algorithms), or offer digital literacy, etc. as Stanford Cyber Policy Center's Nathaniel Persily framed in the "Seven Ds of Disinformation."

[22] Microtargeting plays a critical role in network formation. As Cyber Policy Center program lead and former Facebook chief security officer, Alex Stamos, notes, microtargeted ads are the "tip of the spear…the goal of these ads is not actually to push the message, [but] to build an audience to which the message can then eventually be delivered." Transparency here will be key.

[23] The phenomena of QAnon on the political right, and the #antivax movement on the left, present prime examples. For instance, an internal Facebook report from 2018 noted that 64% of people who joined an extremist group on Facebook only did so because the company's algorithm recommended it to them.

[24] Support for mechanisms to give users themselves more control over what they see have been around since at least since the 1990s. For example, technologies like the Platform for Internet Content Selection were intended to allow users to choose what web content appeared in their browsers. Groups like the E.U. Disinfo Lab note that "users should be empowered to enable and disable algorithmic curation on their feeds and that self-curation features should be provided for by the platforms." As former Google lawyer and current Stanford Cyber Policy Center program lead Daphne Keller notes, "settings on YouTube or Twitter … could include dials and knobs to signal our individual tolerance for violence, nudity, or hateful speech."

[25] There are also many third-party plug-ins that change social media platforms' content according to peoples' needs / preferences, but most require technical expertise to discover and install, and thus user adoption is low.

[26] Criticisms of such controls are widespread. Former Google lawyer and current Stanford professor Daphne Keller rightly notes that "it may be that such tools have only limited promise as a technical

matter, because they depend on accurate content labeling. …[In addition] truly granular user content controls would be an incredible amount of work to implement, [and] still might not work very well." For example, "a user who wanted to block most racial epithets but retain access to rap lyrics, historical documents, and news reporting … could do so only if people or algorithms first correctly identified content in these categories."

27 A related option, putting somewhat less burden on users, has been advocated for by Techdirt's Mike Masnick, Stanford's Francis Fukuyama, Duke's Barak Richman, and others: Says Masnick, "Ideally, Facebook (and others) should open up so that third party tools can provide their own experiences — and then each person could choose the service or filtering setup that they want…", a concept Fukuyama and Richman refer to as "middleware." Keller coins this the Magic API model, and notes that "users might choose a G-rated version of Twitter from Disney or a racial justice-oriented lens on YouTube from a Black Lives Matter-affiliated group."

28 Yet challenges with a Magic API are plentiful. As mentioned, getting users to actually install them presents a significant hurdle. And, as Keller notes, "the technology required to make Magic APIs work would be difficult, perhaps impossible, to build well; that's the 'magic' part." It would be hugely inefficient, as each competitor would have to make enormous investments in analyzing content. Platforms would be unlikely to agree to it and may not be compelled to do so constitutionally. And privacy issues with the creation of new APIs abound, as with Cambridge Analytica (an instance of a third party exploiting Facebook's API). Moreover, if platform regulation is challenging at present, decentralizing platform curation power in this way could make research, oversight, and regulation of these potentially thousands of mini-Facebooks untenable.

29 The original intent was to address pornographic, obscene, and indecent material on the internet.

30 CDA 230 ensures that providers, including social media platforms, can set their own rules to govern the content permitted on their sites, without being considered "publishers" that have editorial control over — and liability for — that content.

31 However, privacy too proves no panacea. As Fukuyama notes, privacy initiatives would likely reduce "only the personalization of news for each individual, not the concentration of editorial power," which Fukuyama and others have referred to as "a loaded weapon sitting on a table." This may seem tolerable under the current (presumably liberal) platform leadership, but "no liberal democracy is content to entrust concentrated political power to individuals based on assumptions about their good intentions." This powerful editorial power must also be addressed.

32 Further complicating privacy as a solution is the concern that "the horse has long since left the barn." If new privacy laws were to prevent new competitors from accumulating and using similar data, such laws would run the risk of simply locking in the advantages of first movers like Google and Facebook. See: https://cyber.fsi.stanford.edu/publication/report-working-group-platform-scale.

33 Facebook earned $69.7 billion from advertising in 2019, more than 98% of its total revenue for the year.

34 As the University of North Carolina's Zeynep Tufekci notes, "ad-based financing means that the companies have an interest in manipulating our attention on behalf of advertisers, instead of letting us connect as we wish." It is this business model that incentivizes algorithms that prioritize engaging — and often misleading — content over quality content.

35 Facebook makes about 20 cents per user, per month in profit. If, instead, just one quarter of Facebook's 1.5 billion users paid $1 per month, it would yield more than $4 billion per year.

36 Such an approach is endorsed by scholars, including Stanford's Lucy Berhnolz and Amherst's Ethan Zuckerman, as well as advocacy organizations like Free Press. Zuckerman posits the following thought experiment: "The digital advertising industry is currently a $333 billion global market. ... Three companies that use intensely surveillant advertising practices — Google, Facebook, and Alibaba — represent $200.3 billion of that market. If we posit a 1% levy on highly surveillant advertising — advertising that incorporates user tracking, combines demographic and psychographic data to create user profiles, or targets using factors other than a user's stated intentions and geography — we can easily posit a $1-2 billion annual fund to support public service digital media." Free Press has proposed a Public Interest Media Endowment to manage the funds generated from such a tax. Nobel Prize-winning economist Paul Romer has similarly proposed "a progressive digital ad revenue tax designed to encourage platform companies to explore models other than surveillant advertising, like subscription-based models."

37 United States per capita funding of public media is less than $4, compared to $30 to $130 per capita invested by countries like Germany, Canada, Sweden, and Great Britain.

38 In 2001, Google adopted the tagline "Don't Be Evil" to describe its corporate code of conduct. In 2015 when Google was restructured within its parent conglomerate, Alphabet, the motto was replaced by "Do the Right Thing," which remains the company's tagline today.

39 Harvard University and the Massachusetts Institute of Technology are jointly offering a new course on the ethics and regulation of artificial intelligence. The University of Texas at Austin introduced the course Ethical Foundations of Computer Science, with the goal of eventually requiring it for all computer science majors. And at Stanford University — the academic core of the tech industry, which, in 1997, founded the now widely critiqued (and reformed) Persuasive Technology Lab — launched, in September 2020, the Ethics, Society and Technology Hub to help coordinate and amplify the teaching, research, and activities on Stanford campus at the intersection of ethics, society, and technology, including courses like the Ethics of Technological Disruption, which, unusually, has been open to both students at Stanford and employees in Silicon Valley.

40 The design and oversight of ethical technologies is complex, in part, due to the often-competing societal demands made of tech companies. In the case of social media, we want privacy and we want transparency and accountability. We want free speech and an accurately informed public. We want a diversity of viewpoints and we want users to primarily consume high-quality content. There are myriad, fundamental tensions in democratic values when it comes to the governance of digital technologies, which need to be more fully explored.

41 The platforms have launched several projects in response to demands for direct data access, including Facebook's Social Science One, launched in 2017, which aimed to provide privacy-protected data on all URLs shared more than 100 times, alongside associated socio-demographic and fact-checking data; Facebook and Google's political ads archives, launched in 2018; and Twitter's archive of state-backed Information Operations, launched that same year.

42 Alternatively, some have suggested that to ensure data access by academics is only used for purposes that serve the public good, the government could instead introduce criminal penalties for researchers who violate data access/usage terms (rather than placing liability for data-protection solely on platforms).

43 Related "citizen science" efforts (where citizens are encouraged to download their own data from Facebook and donate it to academic research) or traditional survey efforts collecting data from users, serve similar ends.

44 Efforts by the Stanford Cyber Policy Center, the Aspen Tech Policy Hub, the Knight Foundation, Craig Newmark Philanthropies, and FakeNewsSci are working to coordinate the field — the first three through convenings and network curation, the latter two primarily through shared listservs. But much more is needed.

45 Cross-platform coordination is essential because the new information environment is interconnected and networked, with disinformation often traversing from relatively fringe sites, like 4chan and Gab, to mainstream social media platforms like Twitter and Facebook, before going mainstream on traditional media platforms. Yet most social media companies operate in siloes, with few forums (or incentives) for companies to share information about emerging disinformation techniques or narratives with their peers or competitors. And even fewer systematically share information with academics or civil society.

[46] This would be helpful in ensuring that academic research and platform interventions are informed by essential information about the needs and experiences of civil society, and/or the intelligence currently not accessible to non-government actors.

[47] GIFCT was formally established in July 2017 by a group of companies focused on disrupting terrorist abuse of member companies' platforms. The original Forum was led by a rotating chair drawn from the founding four companies — Facebook, Microsoft, Twitter, and YouTube — and worked to support knowledge-sharing, technical collaboration, and shared research.

[48] GNI is a Washington, D.C.-based nonprofit founded in 2008 by Google, Yahoo (now Oath), Microsoft, and a small group of nonprofit organizations and universities. Facebook joined GNI in 2013. GNI has no express focus on disinformation / information integrity. Its dual goals focus on "preventing internet censorship by authoritarian governments and protecting the internet privacy rights of individuals."

[49] Many have noted the lack of coordination in this field. The Cyberspace Solarium Commission — established in 2019 as part of the McCain National Defense Authorization Act, to "develop a consensus on a strategic approach to defending the United States in cyberspace against cyber attacks of significant consequences" — has issued a series of reports relevant to disinformation, and has advocated for the establishment of a Social Media Data and Threat Analysis Center (DTAC). Similarly Senator Mark Warner's office has advocated for the requirement of an Interagency Task Force for Countering Asymmetric Threats to Democratic Institutions. Warner notes that "the intelligence and national security communities are not as well-positioned to detect, track, attribute, or counter malicious asymmetric threats to our political system as they should be" and that "programs to detect and protect against information operations are disparately positioned with unclear reporting chains and lack metrics for measuring success."

[50] For example, recent studies reveal that, even in the case of potentially lethal COVID-related disinformation — disinformation that has been debunked by the World Health Organization — weeks later, 59% of it remained up on Twitter, 27% remained on YouTube, and 24% remained on Facebook. Subsequent research found that, of the 41% of the misinformation remaining on Facebook without warning labels, "65% had been debunked by partners of Facebook's very own fact-checking program." This, despite the fact that platforms are heavily incentivized to address COVID-related misinformation given the immediate risks to human life.

[51] Facebook spent more than $16 million in lobbying in 2019 (making it the seventh-largest lobbyist across all industries) and Google surpassed that, with $22 million in lobbying in 2018. It is also worth noting that, with the exception of Blue Cross Blue Shield, Facebook and Google are the only single corporations in the top 10 (all other top-10 lobbyists represent coordinate groups — the Chamber of Commerce, the American Medical Association, etc.)

[52] In addition to the specific remedies discussed above, a range of arguments have been made to strengthen tech oversight more generally: establishing a cybersecurity policy lead in the executive branch; reinstating the Office of Technology Assessment; developing new, independent oversight bodies; and more. Mark MacCarthy of Georgetown University has argued for a dispute resolution program for social media companies: "a transparency and accountability regime enforced by a federal regulatory commission and a dispute resolution system administered by an industry self-regulatory organization aiming to ensure consistency with a company's publicly disclosed content standards." Relatedly, Senators Brian Schatz (D-Hawaii) and John Thune (R-S.D.), in their bipartisan Platform Accountability and Consumer Transparency (PACT) Act, seek to establishes a transparency regime for social media companies that is supervised by the Federal Trade Commission. These interventions would also serve to address challenges far beyond disinformation, providing ongoing governmental capacity to understand and oversee technology companies. But additional research is needed both to better define the problems, and the trade-offs between and within different structures. For example, opponents of greater tech oversight, including most notably the tech companies themselves, rightly note that overseeing "technology" is akin to regulating "innovation," because the services that tech companies themselves offer are so distinct — as mentioned, Apple is a luxury goods manufacturer, Amazon is a delivery service. And the problem becomes even harder with advanced technologies. How does one regulate AI, which is in everything from automatic doors to Alexa?

[53] Most significant among these, by far, was the #StopHateForProfit campaign in the summer of 2020: Organized by a coalition of civil rights groups and other advocacy organizations, the campaign saw more than 1,100 companies worldwide have pull millions of dollars in advertising from the social network, including brands from Ford to Coca-Cola to Unilever. A similar effort by liberal social media activism organization Sleeping Giants has aimed to convince companies to remove advertisements from conservative news outlets. And "whitelisting" efforts by Vodaphone and other corporations have worked stop their own ads appearing alongside sites promoting fake news and hate speech.

[54] Efforts like the Tech Transparency Project are seeking to reveal platform funding of related nonprofit efforts through, for example, the creation of searchable databases tracking corporate contributions to fund-related nonprofits and research.

[55] Source: Foundation Center On