

Seismic Shifts

How Economic, Technological,
and Political Trends are
Challenging Independent
Counter-Election-Disinformation
Initiatives in the United States





The **Center for Democracy & Technology (CDT)** is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.



This report is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Seismic Shifts

**How Economic, Technological,
and Political Trends are
Challenging Independent
Counter-Election-Disinformation
Initiatives in the United States**

**Dean W. Jackson, William T. Adler,
Danielle Dougall, Samir Jain**

With contributions from Asha Allen,
Aliya Bhatia, Alexandra Reeve
Givens, Ari Goldberg, Tim Hoagland,
Emma Llansó, Nathalie Maréchal



Executive Summary

Efforts to protect the integrity of information during elections are threatened by seismic economic, technological, and political shifts.

- Tech sector downsizing has reduced the size of trust and safety teams; new platforms and new technologies like generative AI are making counter-disinformation more challenging.
- A coordinated political assault on election integrity threatens the capacity and, in some cases, the safety of independent counter-disinformation researchers and advocates—those who are unaffiliated with either platforms or government.
- This report draws on interviews with 31 individuals (including current and former tech company employees and representatives from independent research and advocacy initiatives) about their experience responding to election disinformation and the growing challenges they face.

The election integrity initiatives examined for this report reflect a variety of approaches.

- Some consider themselves primarily researchers while others do research to inform advocacy—though almost all include an element of rapid response through methods like counter-messaging to voters or cooperation with election officials.
- Their interaction with government agencies varies from routine meetings to strict policies of non-communication.
- Their relationships with social media platforms similarly range from formal partnerships to distanced criticism.

For independent counter-election-disinformation initiatives, partnerships with platforms provide important benefits but also raise concerns about sustainability and extractive labor.

- Sometimes, initiatives bring cultural and linguistic fluency that platform staff lack and can track harmful narratives that may otherwise go unnoticed or unaddressed by platforms.
- Platform staff are also keenly aware that input from outside experts helps legitimize decisions about content and integrity.
- But some independent professionals are wary of providing “free labor to multi-billion dollar corporations,” calling it an “extractive” but “unfortunately necessary way to reduce harm.”

The 2020 election, tech sector layoffs, and other recent events have called into question the ability of counter-election-disinformation initiatives to influence platform content moderation.

- Interview subjects detailed that platforms were frustratingly inconsistent and sometimes unresponsive before widespread tech sector layoffs. The situation is worse now.
- Likewise, independent researchers have limited insight into digital threats and trends because most platforms are opaque and offer little access to data crucial to answering key questions.
- Generative artificial intelligence poses new potential risks related to election disinformation. Rather than jumping to conclusions, stakeholders should methodically consider the highest potential dangers and most appropriate responses.

As prominent politicians and a significant portion of the electorate continue to deny the outcome of the 2020 election, disinformation researchers have found themselves under attack.

- Independent researchers increasingly face hostile campaigns from partisan media and legal, digital, and sometimes physical harassment—illustrated by Congressional subpoenas to leading figures in the field.
- The chilling effect alone may drive young professionals from the field, make it more difficult to secure funding, and dissuade government officials from engaging with counter-disinformation efforts—especially with the possibility that a pending court case, *Missouri v. Biden*, will result in permanent restraints on government communications with platforms or researchers.

In this environment, independent counter-election-disinformation initiatives are reconsidering their approaches.

- Many initiatives are pivoting harder into other strategies like counter-messaging, assistance to targeted election officials, and policy advocacy.
- Meanwhile, many online trust and safety outcomes are as bad if not worse than in 2016. The 2024 election is likely to be the most vulnerable environment for political disinformation that the United States has seen in eight years.

Recommendations

Independent counter-disinformation initiatives should take steps in the short-to-medium term to weather the storm and mitigate harm.

- Funders, research institutions, and nonprofits should create shared resources and practices for researchers under attack. These might include pools for legal defense, cybersecurity assistance, and proactively developed communications plans for responding to coordinated attacks.
- Counter-election-disinformation should pivot to year-round harm reduction strategies like pre-bunking, training for election officials, and advocacy efforts. False narratives about election fraud persistently impact voting rights between election cycles, so this work should receive consistent support.
- Advocates should focus less on individual pieces of content and more on mitigating the impact of disinformation “superspreaders.” These are a proven force multiplier for mis- and disinformation—relatively few individuals are responsible for a great deal of false, viral content.

In the medium-to-long term, election integrity initiatives should widen the aperture for advocacy—relying less on unstable partnerships with platforms and the federal government to other stakeholders and non-digital threats to elections.

- Researchers, donors, and advocates should treat election disinformation as part of a larger, institutional problem by supporting reforms to the electoral process and law. Electoral systems like ranked choice voting and primary reform may reduce incentives for disinformation.
- Advocates and their donors should increase the resources spent on advocacy to select state governments around relevant issues like security for election workers and researcher access to data.

Government can also take steps to promote public confidence in election integrity and counter-disinformation efforts and, in conjunction with other stakeholders, do more to promote online trust and safety while respecting freedom of expression.

- Government and other institutions should promote and make use of former trust & safety staffers' talent by hiring them and encouraging the profession to develop norms, standards, and field-building opportunities comparable to related industries like cybersecurity.
- Governments should clarify and be more transparent about their role in responding to election disinformation—especially in the aftermath of the injunction issued in the case of *Missouri v. Biden*. They could explicitly set boundaries and transparency requirements around federal government communications with social media platforms and independent researchers.

Platforms should improve both capacity and process for protecting elections from digital disinformation.

- Platforms should reinvest in trust and safety teams as soon as possible, focusing especially on civil rights specialists who can shape content moderation policy and practice.
- Platforms should recommit to policies and practices that combat election disinformation, and respond to claims of censorship and bias by adhering to principles such as the Santa Clara Principles on Transparency and Accountability in Content Moderation.
- Platforms should designate consistent points of contact for civil society. The departure of key personnel shows the limits of personalized relationships and has been a persistent problem for independent researchers.
- Platforms should of their own accord increase transparency around their communications with government agencies.
- Platforms should expand researcher access to platform data—and lawmakers should consider supporting that expansion through legislation like the Platform Accountability and Transparency Act. The public deserves to know more about the impact of social media on society.

Stakeholders should be attentive to the potential risks generative AI presents for election integrity.

- Stakeholders from all sectors should methodically and carefully parse risks and predictions related to AI—taking stock of available evidence, key questions to which answers are not yet known, and which risks merit priority response.
- Potential responses might lie in better systems for detecting and labeling AI-generated content, changes to political advertising law, better consumer protection rules, and public education.



Contents

Executive Summary	4
Recommendations	7
Introduction	12
Scope & Methodology	14
Models for Independent Election Integrity Initiatives	16
Examples of Counter-Election-Disinformation Initiatives	17
Comparing and Contrasting Counter-Election-Disinformation Initiatives	34
Why Are Partnerships With Independent Initiatives Valuable?	39
Challenges for Independent Counter-Election-Disinformation Initiatives	45
Partnerships and Approaches Are Too Time-Bounded	46
Platforms Are Becoming Less Responsive	48
Lack of External Access to Platform Data Causes Problems for Both Platforms and Researchers	57
New, “Alternative” Platforms Complicate Trust & Safety	61
Generative AI Brings New Risks	62
Political Retaliation is a Growing Threat to Disinformation Research	65
Funding Challenges Cloud the Horizon	78



Adapting for 2024 and Beyond	80
What do Disinformation Researchers Need in a Hostile Environment?	81
Having Soured on Platform Partnerships, Many Initiatives Are Exploring Other Approaches	84
Recommendations	89
Short-to-Medium Term Steps to Protect Researchers & Mitigate Harm	91
Medium-to-Long Term Strategic Shifts for Election Integrity & Advocacy	94
Government Steps to Promote Trust & Safety and Public Confidence in Elections	100
Methodical Consideration of Generative AI and its Potential Risks	102
Conclusion	104
Appendix: List of Interviews	106



01

Introduction

EFFORTS TO PROTECT THE INTEGRITY OF INFORMATION DURING ELECTIONS STAND AT A CROSSROADS. They have exited a start-up phase bookended by two U.S. presidential elections: the first in 2016, when Russian efforts to divide the public and influence the election's outcome caused international shock and scandal, and the second in 2020, when the losing candidate used false claims of fraud to incite violence and disrupt the transfer of power.

While not all disinformation is digital, over the last seven years academia and civil society have paid a great deal of attention to the role of social media and technology companies in the spread of false claims about elections and their responsibility to moderate such claims. Working outside and independently of platforms, researchers strive to better understand the interplay between the internet and politics, investigators search for efforts to manipulate public opinion, and advocacy groups endeavor to reform the technology sector and protect vulnerable demographics from harm.

Even well-resourced monitoring efforts struggle to keep pace with the torrent of election falsehoods online.

The landscape for their work has undergone seismic shifts since the U.S. 2020 presidential election and the subsequent midterm elections in 2022. The number of digital platforms continues to increase, expanding the surface that researchers must monitor as bad actors use multiple platforms to spread their messages. Many new platforms (and, increasingly, X, the platform formerly known as Twitter, under Elon Musk’s ownership) are ideologically hostile, or at least ambivalent, to the principles of content moderation under which the largest platforms have responded to mis- and disinformation in the past. Some of these have become gathering places for extremists and conspiracists who engage in activity that is prohibited elsewhere. Misleading content in languages other than English continues to circulate relatively unchecked. The advent of generative AI threatens to create an “infinite” supply of disinformation.¹

Even well-resourced monitoring efforts struggle to keep pace with the torrent of election falsehoods online. Despite the magnitude of the challenge, platforms that previously invested heavily in election integrity have cut staff and budgets for trust and safety work as a result of diminishing political pressure and workforce reduction across the technology sector.

Efforts to protect U.S. elections from disinformation also face legal challenges that make their future uncertain. In *Missouri v. Biden*, an ongoing case brought by the Attorneys General of Missouri and Louisiana alleging government imposition of social media censorship during the COVID-19 pandemic and the 2020 election, a federal judge issued an injunction, subsequently narrowed by the Fifth Circuit and stayed at the time of this writing, restricting the ability of significant parts of the government from interacting with either platforms about social media content moderation. The lower court injunction had enjoined communications with prominent disinformation researchers, which an amicus brief filed by Stanford University in July 2023 describes as “based upon conjecture, misunderstandings, and, in at least two instances, invented

1 DiResta, R. (2020, September 20). [The Supply of Disinformation Will Soon Be Infinite](https://perma.cc/Q3SK-LXZ8). *The Atlantic*. [perma.cc/Q3SK-LXZ8]

quotations never uttered” by the individuals in question.² While the outcome of *Missouri v. Biden* is uncertain, the mere existence of the injunction has already chilled government counter-disinformation efforts and engagement with other stakeholders.

Last but not least, the work of independent election disinformation researchers has become more difficult as a result of surging harassment, legal threats, and political reprisals. Some have gone so far as to call Elon Musk’s release of the “Twitter Files” and the subsequent Congressional hearings in the House Select Subcommittee on Weaponization of the Federal Government an organized attack on their profession, saying that “The ‘weaponization’ committee is being weaponized against us.”³

Scope & Methodology

This report explores the role of independent election integrity initiatives—those outside of government or technology companies—in protecting elections from digital disinformation and its consequences. It also describes how new obstacles and risks associated with election disinformation monitoring are changing how these initiatives conceptualize their goals and their relationships with tech companies. It is primarily based on interviews with relevant individuals from various initiatives and platform staff familiar with their work. It examines the strengths and weaknesses of their divergent approaches and discusses their challenges and subsequent adaptation. Most of the individuals interviewed for this report declined to be named so that they could speak freely. In those instances, we identify them by their relevant employer (or in some cases, former employer) rather than by their name. Many individuals declined to be interviewed

2 [Brief of amici curiae Stanford University, Alex Stamos, and Renée DiResta in Support of Appellants, *Missouri v. Biden* \(2023\). \[perma.cc/63BD-HPUV\]](#)

3 [Bernstein, A. \(2023, March 22\). Republican Rep. Jim Jordan Issues Sweeping Information Requests to Universities Researching Disinformation. *ProPublica*. \[perma.cc/3T2Y-EHCC\]](#)

at all, either explicitly or presumably because of the fear of legal or political consequences. This report also draws on research reports, journalistic coverage, and public statements from platforms to corroborate claims.

This report mostly focuses on the United States, where stalled regulatory efforts such as the Platform Accountability and Transparency Act and growing political pushback present obstacles even as electoral threats remain significant.⁴ In Europe and elsewhere, governments have taken more aggressive regulatory stances and backlash to information integrity efforts have not yet reached critical mass. Some insights from these contexts are included for comparative purposes.

Finally, a note on terminology: this report uses the term “disinformation” largely as an umbrella term for related but distinct claims about elections. The most common definition applies here: purposeful attempts to mislead. The Election Integrity Partnership refers to “rumors,” which can turn out to be true but may not be verified at the time of the claim. Others use “disinformation” to refer to harmful narratives or content that may be misleading, divisive, or incendiary but are subjective and non-falsifiable. Some also include propaganda from ideological extremists. For the sake of simplicity and to center its analysis on the most pernicious forms of the problem, this report uses the phrase “disinformation” throughout.

4 [Platform Accountability and Transparency Act](#), S. 5339, 115th Cong. (Introduced 2023, December 21). [perma.cc/73UH-2MAS]

02

Models for Independent Election Integrity Initiatives

Independent election integrity initiatives have a few common features, but many differences. What they share is their independence—though some communicate with platforms or government agencies and officials, they exist outside of those organizations. Typically, they are housed within universities or nonprofit organizations, although there are examples of other models (such as private, for-profit investigative firms, and research consultancies). They also share a broad objective: the protection of free and fair elections from false claims, especially when those claims are made purposefully and knowingly to influence, overturn, or discredit election results.

Their approaches, however, differ. Many liaise with platforms, law enforcement, or election officials for the purposes of providing situational awareness of threats to the election process. Others eschew these relationships as unproductive or compromising. Some inform law enforcement when they find evidence of a risk of imminent harm or foreign interference. Despite public allegations, no credible evidence has emerged that these initiatives used contacts with law enforcement for partisan purposes⁵ or that law enforcement used findings from these initiatives to pressure social media companies to take action.

5 Public statements from researchers cited throughout this report and the above-cited [amicus brief](#) submitted by Stanford in *Missouri v. Biden* offer in-depth refutations of several key claims regarding this allegation. [perma.cc/63BD-HPUV]

There are differences in size and make-up of these initiatives: Some are small, comprising only a few closely coordinating organizations or individuals. Others involve dozens of institutions that belong to a central umbrella organization. Many are networks of professionals, but some incorporate volunteers.

While representatives from several initiatives were interviewed for this report, three received most of our attention: the Election Integrity Partnership (EIP), Common Cause, and the Disinfo Defense League (DDL). These efforts take differing approaches and have a wide vantage point from which to observe their field, making them useful primary case studies. Table 1 summarizes several examples of these initiatives.

Examples of Counter-Election-Disinformation Initiatives

The Election Integrity Partnership

The Election Integrity Partnership,⁶ founded in summer 2020, combined the efforts of four established disinformation research centers: the Stanford Internet Observatory, the University of Washington Center for an Informed Public, a think tank center at the Atlantic Council called the Digital Forensic Research Lab (DFRLab), and a private open-source investigative firm called Graphika (though the latter two did not participate in 2022). These four teams created a joint ticketing system to identify, track, and report instances of election mis- and disinformation to election integrity teams at platforms. A “ticket” could be as granular as a single social media post, or it might capture an entire website dedicated to election falsehoods or a meme that spread across several platforms. According to its final report, many of the EIP’s tickets in 2020 were related to attempts to delegitimize various parts of the election process such as vote counting, voting machines, and mail-in voting.

6 Election Integrity Partnership. (n.d.). [The 2020 Election Integrity Partnership](https://www.electionintegritypartnership.org/). [perma.cc/AT5G-QKBU]



Table 1. Examples of Counter-Election-Disinformation Initiatives.

For all rows except the EIP, the source of this information came from interviews between the author and organization staff. There are a number of ways organizations might communicate or work with government agencies or platforms. This chart is meant to summarize routine or frequent interactions related to election content. As such, it may not reflect all current and historical relationships or interactions between stakeholders. In many cases, interviewees were hesitant to characterize these relationships in detail due to security and political considerations.

Like many organizations in the advocacy space, Common Cause is, legally speaking, two organizations: A 501(c)(4) called Common Cause and a 501(c)(3) called Common Cause Education Fund. Advocacy activities are undertaken by the (c)(4); the (c)(3) conducts Common Causes’s disinformation monitoring.

ISD is listed as having “mixed” communications about content with platforms because of conflicting information. In an interview, a researcher affiliated with ISD said that partnering with platforms is not a major part of their counter-disinformation work in the United States. However, interviewees from one major platform said their employer does have a relationship with ISD, suggesting that interaction between the two organizations is limited in scope or inconsistent across the organization.

Finally, in an interview, EDMO staff said they interact with platforms through public forums but do not partner with them directly or formally.

Election Integrity Partnership (EIP)

Description	
<p>The EIP initially consisted of four organizations that worked together to monitor “attempts to suppress voting, reduce participation, confuse voters, or delegitimize election results without evidence” during and after the 2020 election and better equip platforms, election officials, government agencies, and civil society to respond. In 2020, its four members were the Atlantic Council’s Digital Forensic Research Lab (DFRLab), Graphika, The University of Washington Center for an Informed Public, and the Stanford Internet Observatory (SIO). In 2022, The University of Washington and the SIO were core conveners, and the DFRLab, Graphika, and the National Conference on Citizenship acted as partners.</p>	
Approach	Communicates about content with:
<p>The EIP created a shared ticketing system that allowed researchers from four organizations working in coalition to identify, flag, and categorize different types of election disinformation and provide their findings to stakeholders in social media companies and law enforcement. The EIP did not consider it a goal to change platform or government policy and did not advocate for specific actions to be taken against users.</p>	<p>Government</p> <p>In 2020, the EIP described its mission as “supporting real-time information exchange between the research community, election officials, government agencies, civil society organizations, and social media platforms.” One avenue for this was its participation in the EI-ISAC, a cybersecurity center supporting election officials. It did not participate in a similar exercise in 2022.</p>
	<p>Platforms</p> <p>Yes</p>

Common Cause

Description	
<p>Common Cause is a civil society organization committed to fair elections and voting rights. Since 2016, it has monitored disinformation before, during, and after elections to enable better rapid response by platforms, government, and civil society—including more effective content moderation by platforms and counter-messaging by government officials and civil society.</p>	
Approach	Communicates about content with:
<p>In 2020 and 2022, Common Cause ran a network of protection election volunteers who monitored social media platforms for instances of election disinformation or intimidation against election officials and reported that content to platforms when they believed it violated company terms of service. Volunteers were also equipped with counter-messages meant to pre- and debunk common disinformation tropes during the election period.</p>	<p>Government</p> <p>Common Cause national staff and state chapters both sometimes communicate with state election officials. Common Cause has also submitted Congressional testimony on disinformation.</p> <p>Platforms</p> <p>Yes</p>

Disinfo Defense League (DDL)

Description	
<p>The DDL was established by Media Democracy Fund to help empower and coordinate civil society efforts to respond to racialized disinformation affecting marginalized communities. It is a closed network that does not publicly list its members and has a very small dedicated staff. The results of its work are primarily research insights, counter-messaging approaches, and policy recommendations.</p>	
Approach	Communicates about content with:
<p>In 2020 and 2022, the DDL hosted training webinars and expert briefings for its members; it also produced topline reports of prominent and dangerous disinformation narratives and helped its members produce counter-messages to them. Since 2022, it has been shifting to producing quarterly in-depth reports looking at the impact of disinformation narratives on specific communities.</p>	<p>Government</p> <p>No</p> <p>Platforms</p> <p>No</p>

Institute for Strategic Dialogue (ISD)

Description	
<p>ISD conducts programs and analysis responding to extremism “in all its forms.” It is a global organization headquartered in London but with offices in Washington, DC; Berlin; Amman; Nairobi; and Paris.</p>	
Approach	Communicates about content with:
<p>In the United States, ISD’s work on election disinformation mostly consists of open-source intelligence gathering on the activities of extremist organizations and movements. ISD works with election officials when the situation warrants to protect the security and integrity of polling places. In situations with a high risk of offline violence, they may contact law enforcement.</p>	<p>Government</p> <p>When online threats against election officials appear credible, ISD may work with those officials and with law enforcement to respond.</p>
	<p>Platforms</p> <p>Mixed</p>

Anti-Defamation League Center on Extremism

Description	
<p>The ADL Center on Extremism tracks “extremist trends, ideologies, and groups” across the ideological spectrum. It produces a map of hate and extremist activity across the United States and conducts open-source investigations of extremist activity online.</p>	
Approach	Communicates about content with:
<p>The ADL Center on Extremism monitors known hate and extremist groups online to better understand their activities and how they can be countered. In situations where offline violence is likely, staff may communicate with law enforcement.</p>	<p>Government</p> <p>When extremist rhetoric online contains credible threats, ADL may report it to law enforcement.</p>
	<p>Platforms</p> <p>Yes</p>

Carter Center

Description	
<p>The Carter Center is an international NGO with a long history of international election monitoring. It does not monitor U.S. elections.</p>	
Approach	Communicates about content with:
<p>The Carter Center monitors social media and digital disinformation in the lead-up to elections using teams of local contractors. It does not typically work on platform policy but did produce a report on “The Big Lie and Big Tech” in 2021.</p>	<p>Government</p> <p>In especially concerning situations, particularly those related to imminent threats, the Carter Center may contact election authorities.</p>
	<p>Platforms</p> <p>No</p>

European Digital Media Observatory (EDMO)

Description	
<p>EDMO involves a large number of partner organizations across the European Union, organized into national and regional hubs covering every EU Member State. It evolved out of a previous initiative as a more research-oriented offshoot.</p>	
Approach	Communicates about content with:
<p>EDMO and its hubs work together to understand online threats to democracy and inform the EU policy-making processes. They do not correspond with platforms or law enforcement directly although they may present their findings in open forums where those organizations access them.</p>	<p>Government</p> <p>No</p>
	<p>Platforms</p> <p>No</p>

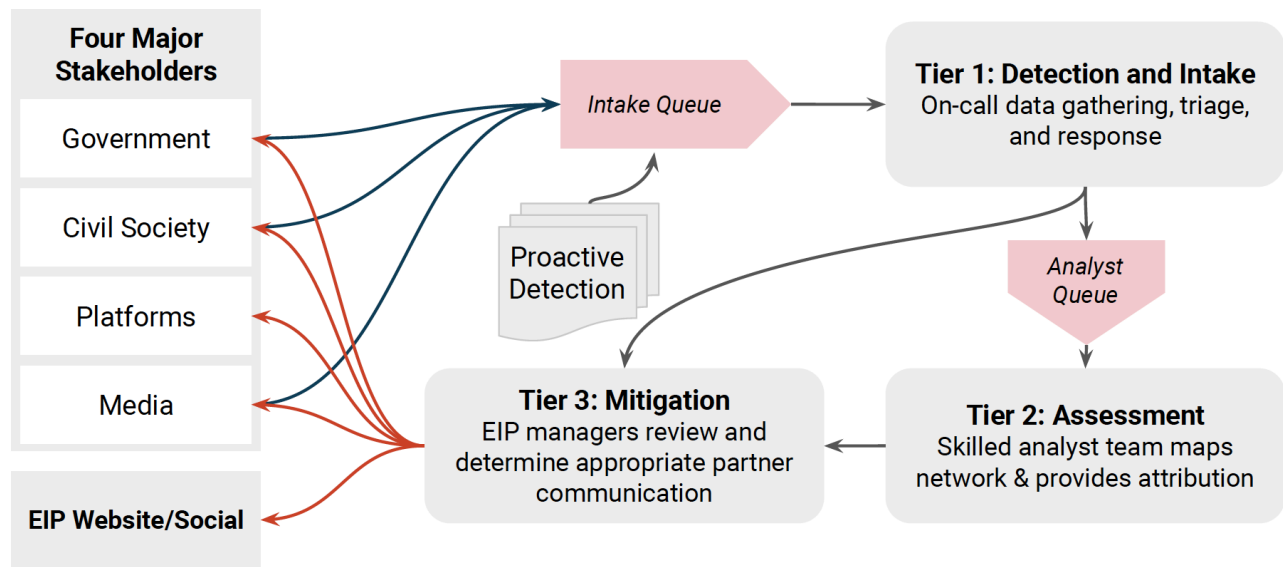


Figure 1. EIP's Internal Workflow.

Chart taken from the EIP final report showing the EIP's internal workflow. Tickets are filed and then move through the process in the directions indicated by arrows.

Source: *The Election Integrity Partnership*. (2021, June 15). *The Long Fuse: Misinformation and the 2020 Election*.

The EIP worked closely with social media platforms throughout the ticketing process—so closely, in fact, that representatives from Facebook, Instagram, Google, YouTube, Twitter, TikTok, Reddit, Nextdoor, Discord, and Pinterest were “onboarded” into the process so that if the EIP believed a ticket included a terms of service violation, those corporate representatives could be added directly to the ticket.

The EIP focused specifically on rumors that might contribute to interference in election procedures, suppress voter participation, falsely allege fraud, and delegitimize election results. It described its goals during the 2020 election as “identifying misinformation before it goes viral”; “sharing clear and accurate counter-messaging”; and “increasing transparency into what happened during the 2020 elections.”⁷ Based on past elections and expectations for 2020, the EIP predicted several misleading narratives in advance and worked with journalists to “pre-bunk” them, aiming to essentially inoculate the public.⁸

7 The Election Integrity Partnership. (2021, June 15). [The Long Fuse: Misinformation and the 2020 Election](https://perma.cc/E3YK-D239). [perma.cc/E3YK-D239]

8 Research suggests inoculation and pre-bunking can be effective at blunting disinformation's influence, but they can be difficult because they must be done in advance. Consider: Garcia, L., Shane, T. (2021, June 29). [A guide to prebunking: a promising way to inoculate against misinformation](https://perma.cc/K77Q-L65T). *First Draft News*. [perma.cc/K77Q-L65T]; Roozenbeek, J., van der Linden, S., & Nygren, T. (2020, February 3). [Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures](https://perma.cc/EBE6-YY3E). *Misinformation Review*. [perma.cc/EBE6-YY3E]



Figure 2. Example EIP Ticket.

Image taken from the EIP final report. URLs and names of staff have been redacted by EIP.

Source: *The Election Integrity Partnership*. (2021, June 15). *The Long Fuse: Misinformation and the 2020 Election*.

SHARPIEGATE

[Redacted] raised this on 04/Nov/20 10:25 AM [Hide details](#)

Description

#Sharpiegate is trending on twitter after allegations that voters were forced to use sharpie Maricopa County in Arizona and that the sharpie was intentionally meant to make votes ambiguous so to sway the election.

This is not true. The ballots are designed such that sharpie ink will not compromise the selection.

This has spread to a variety of different states across Twitter, FB, TikTok, and Youtube, we will use this ticket to try and consolidate all the content. While the primary reports have come from Arizona, similar claims of felt-tipped markers being illegally used to sway election outcomes have been made across Chicago, IL and Shasta County, CA.

URLs

<https://twitter.com/> [Redacted]

<https://twitter.com/> [Redacted]

<https://twitter.com/> [Redacted]

<https://twitter.com/> [Redacted]

<https://twitter.com/> [Redacted]

<https://twitter.com/> [Redacted]

<https://vm.tiktok.co> [Redacted]

<https://www.instagram> [Redacted]

<https://www.youtub> [Redacted]

Status

IN REVIEW

Request type

EIP Report

Shared with

[Redacted]

- TikTok
- Facebook
- EI-ISAC
- Google
- Twitter
- [+](#) Share

The EIP worked to track instances of election rumors before and after the election across fifteen platforms. In addition to those onboarded to the ticketing process, the list includes sites sometimes described as “alt-tech” platforms for their ideological opposition to content moderation as practiced by mainstream platforms.⁹ These were monitored by EIP but not partnered with (in many cases because they had no relevant content moderation policies).

⁹ Newton, C. (2021, July 6). [Conservative social networks keep making the same mistake](https://www.theverge.com/2021/7/6/23011111/conservative-social-networks-keep-making-the-same-mistake). *The Verge*. [perma.cc/QJU9-BZYS]

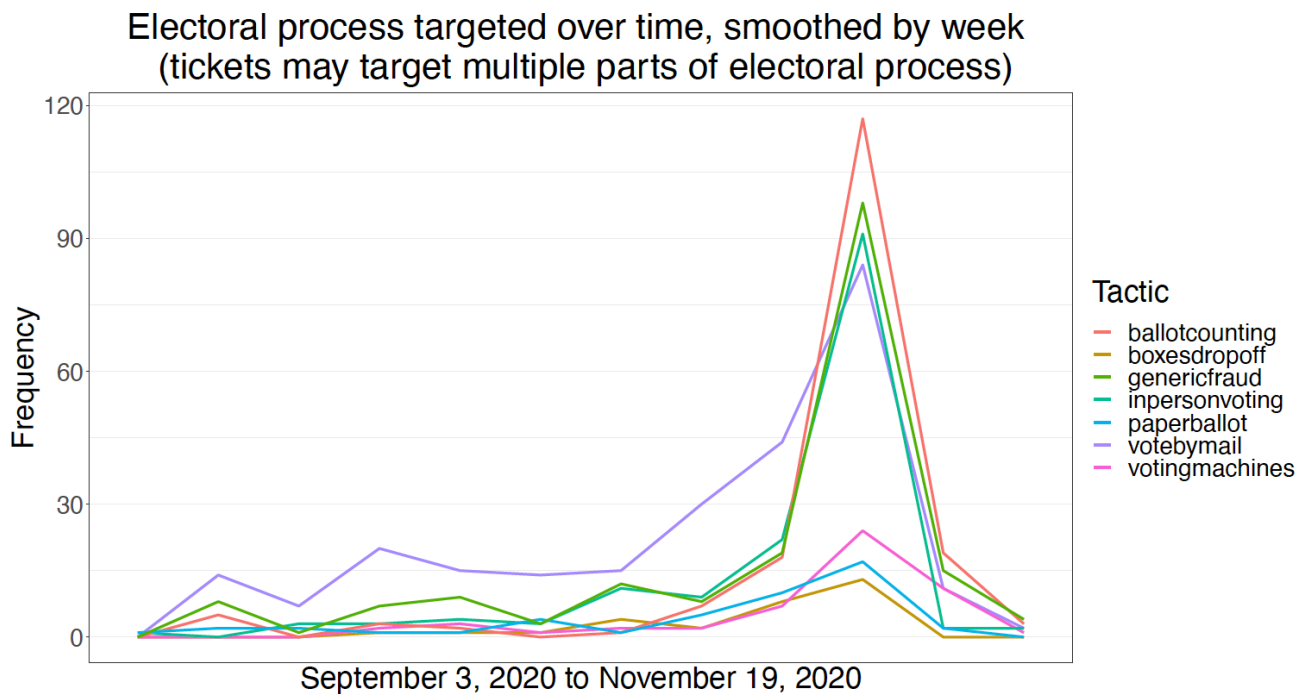


Figure 3. EIP Tickets Sorted by Relevance to Election Process Over Time.

Image taken from the EIP final report. The spike in tickets corresponds to election day 2020.

Source: *The Election Integrity Partnership*. (2021, June 15). *The Long Fuse: Misinformation and the 2020 Election*.

The EIP also engaged with government stakeholders, who collaborated with one another, the private sector, and third-party researchers through mechanisms like the Election Infrastructure Information Sharing and Analysis Center (EI-ISAC), run by an independent nonprofit called the Center for Internet Security. According to the EIP's final report, the EI-ISAC "served as a singular conduit for election officials to report false or misleading information to platforms." This allowed government officials to flag digital threats to the election to both platforms and third-party partners and to receive situational analysis and potential counter-messaging guidance from researchers like those in the EIP. Despite unfounded allegations to the contrary, members of the EIP have issued several emphatic statements that it does not make content moderation decisions for platforms and that it did not pass reports to social media companies at the request of the Department of Homeland Security or the Cybersecurity and Infrastructure Security Agency.¹⁰

10 Election Integrity Partnership. (2022, October 5). [A Statement from the Election Integrity Partnership](https://perma.cc/S6RV-XZWE). [perma.cc/S6RV-XZWE]

Common Cause

Common Cause is a co-lead of the non-partisan Election Protection Coalition, which unites over 300 national, state, and local partners who work year-round to protect voting rights.

Other networks are larger than the EIP. Common Cause is a co-lead of the non-partisan Election Protection Coalition, which unites over 300 national, state, and local partners who work year-round to protect voting rights.¹¹ Common Cause also coordinates with volunteers to monitor platforms and to produce and disseminate counter-messages to the disinformation narratives they find there. In an interview, one Common Cause employee cited their work to prepare the public for a “red mirage,” the early appearance of

Republican leads before mail-in ballots are counted.



In recent elections, Republicans have been more likely to vote in person than by mail; concerned that gains by Democrats as mail-in ballots were counted would empower false claims of voter fraud, Common Cause worked to push out messages addressing this issue early in the election.¹²

Common Cause also uses the information collected by volunteers and on-staff analysts to work with social media companies “to remove the content we find and hold them accountable to strengthen their policies.”¹³ Common Cause provided several examples of tweets submitted to platforms this way; some of them are included in figures 4,¹⁴ 5,¹⁵ 6,¹⁶ and 7.¹⁷

11 Common Cause has given Congressional testimony on these issues. See: Getachew, Y. (2022, June 20). Testimony on “A Growing Threat: How Disinformation Damages American Democracy” before the U.S. House of Representatives, Committee on House Administration, Subcommittee on Elections. [perma.cc/69LW-YTVQ]

12 For examples of Common Cause’s messaging, see: Common Cause. (n.d.). Election Night is Not Results Night. [perma.cc/TU3F-ZHWC]. In interviews, Common Cause staff shared that outside polling suggests the public was largely aware there would be delays in election results in 2020, indicating a successful messaging effort.

13 Common Cause. (n.d.). Stopping Cyber Suppression. [perma.cc/D9M8-CUVM]

14  [RealRobert](#)  [@Real_RobN](#). (2022, September 12). [Tweet]. Twitter. [perma.cc/AW93-CBNF]

15 Kelly Loeffler [[@Kloeffler](#)]. (2022, December 6). [Tweet]. Twitter. [perma.cc/GW2Q-5QU3]

16 Marjorie Taylor Greene [[@mtgreenee](#)]. (2022, November 29). [Tweet]. Twitter. [perma.cc/44JA-WESR]

17 Sassy Madeline Maga [[@MadelineYMaga](#)]. (2022, October 25). [Tweet]. Twitter. [perma.cc/BZL5-6VDW]



Figure 4. Tweet from 2022 alleging election fraud in 2020.

Screenshot taken from link provided by Common Cause.

Source: RealRobert [[@RealRobN](#)], (2022, September 12), [Tweet]. Twitter. [[perma.cc/AW93-CBNF](#)].



Figure 5. Former U.S. Senator Kelly Loeffler (R-Georgia) claiming that armed groups of Black Panthers were patrolling polling places in 2022.

Screenshot taken from link provided by Common Cause. Politifact found “no evidence” of this. [[perma.cc/74KV-ZVHF](#)]

Source: Kelly Loeffler [[@Kloeffler](#)], (2022, December 6), [Tweet]. Twitter. [[perma.cc/GW2Q-5QU3](#)]



Common Cause volunteers agree to act in a nonpartisan manner while providing citizens accurate information about the voting process and monitoring public social media channels for election-



Figure 6. Tweet from U.S. Representative Marjorie Taylor Greene (R-Georgia) denying the results of the Arizona gubernatorial election.

Screenshot taken from link provided by Common Cause. Vocal election denier Kari Lake lost the race for Arizona governor in 2022.

Source: Marjorie Taylor Greene [[@mtgreenee](#)]. (2022, November 29). [Tweet]. Twitter. [[perma.cc/44JA-WESR](#)].

Marjorie Taylor Greene @mtgreenee · Nov 29, 2022

I'm proud of @KariLake for fighting to protect the people of Arizona's votes.

Without secure elections, we are no better than third world countries and have lost our freedoms.

AZ SOS Katie Hobbs refused to debate Kari bc she knew it was rigged and didn't have to.

1/2

1,037 3,722 20.4K

Marjorie Taylor Greene @mtgreenee

States with massive mail in ballots, ballot harvesting, no voter ID, machines no one trust, and continued vote counting that turns Election Day into election month are a joke.

Call me all the petty names you want, I could care less.

Kari Lake didn't lose her election.

Readers added context they thought people might want to know

Katie Hobbs defeated Kari Lake in the Arizona gubernatorial election. The race has been called by all major publications. As of 11/29/22 Cochise County is the only remaining county in Arizona to not certify its election results.

[nytimes.com/interactive/20...](#)

[azcentral.com/story/news/pol...](#)

Do you find this helpful? **Rate it**

related mis- and disinformation.¹⁸ An extension of this work is Common Cause's Stopping Cyber Suppression program, a disinformation monitoring and reporting initiative designed in part to educate voters on what to do if they encounter voting-related disinformation online.¹⁹

18 Common Cause. (n.d.). [Nonpartisan Agreement](#). [[perma.cc/H35R-K2VW](#)]

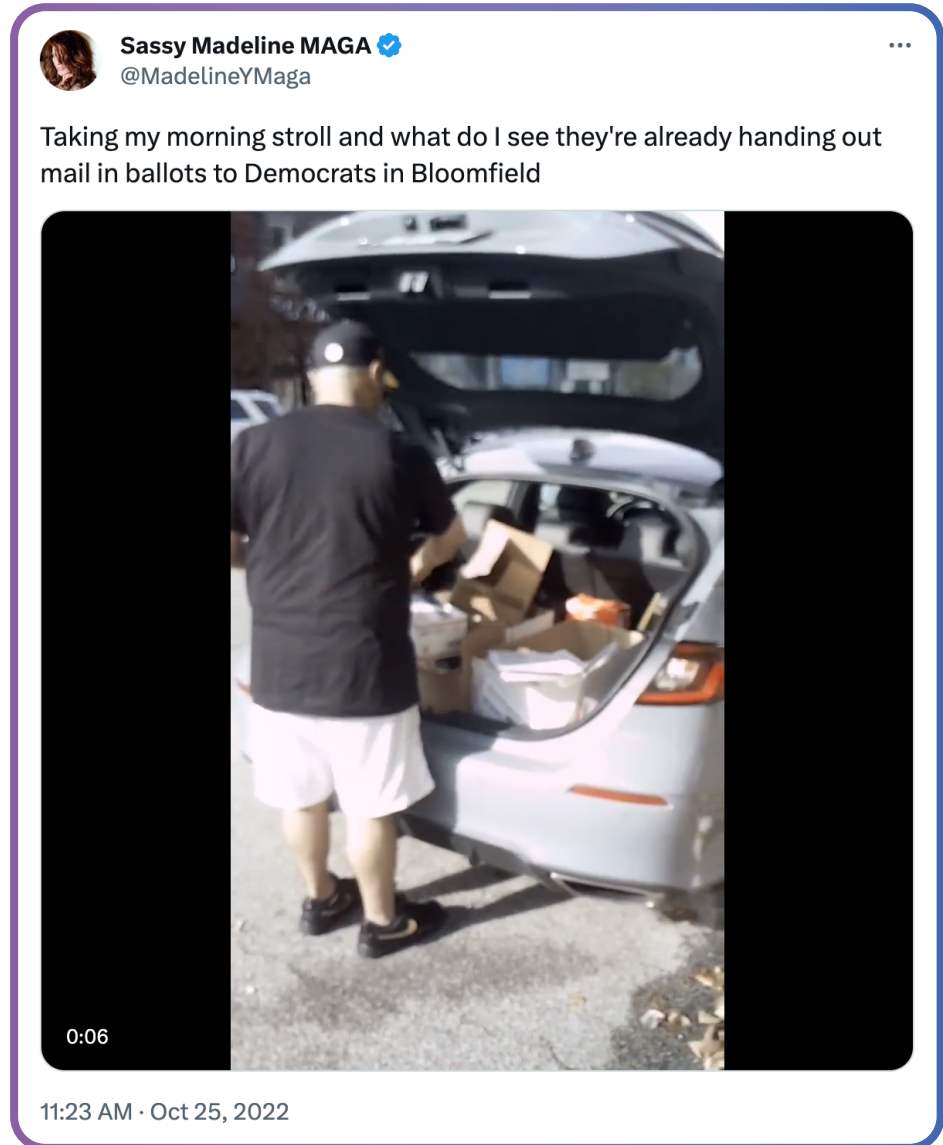
19 Common Cause. (n.d.). [Stopping Cyber Suppression](#). [[perma.cc/D9M8-CUVM](#)]



Figure 7. October 2022 tweet from an individual confronting canvassers and accusing them of ballot fraud.

Screenshot taken from link provided by Common Cause.

Source: Sassy Madeline Maga [[@MadelineYMaga](#)]. (2022, October 25). [Tweet]. Twitter. [perma.cc/BZL5-6VDW]



A researcher affiliated with Common Cause said in an interview that volunteers are especially valuable for covering niche, locally relevant social media spaces like NextDoor and closed Facebook groups; Common Cause sees this as a distinct advantage of its approach, as it improves their visibility across the internet by helping them understand what is cross-posted between platforms. Volunteers can also capture important offline events, like public statements from rogue election officials making false claims of fraud. However, Common Cause emphasizes that volunteers are not asked to monitor the most dangerous private channels where domestic extremists communicate.

The Disinfo Defense League

Not all third-party counter-election-disinformation coalitions consider it their mission to provide actionable information to platform trust and safety teams. The DDL began in the summer of 2020, connects more than 230 member organizations, and largely eschews relationships with platforms (though many of its members do have their own working relationships with platform integrity teams).²⁰

Instead, the DDL focuses on combating “racialized disinformation,” which it defines as “false or intentionally misleading communication or propaganda—typically about racial or social justice issues—that are strategically created and distributed through online media to deceive or manipulate the public for the purposes of achieving profit, political gain, and/or sustaining white supremacy.” In 2020, this included narratives about the election, especially those aimed at suppressing the vote of marginalized communities.

DDL staff stressed that its focus on race and its diverse membership gives it access to lived experience that can be invaluable for understanding disinformation’s impact on specific communities.

In interviews, DDL staff stressed that its focus on race and its diverse membership gives it access to lived experience that can be invaluable for understanding disinformation’s impact on specific communities. DDL has primarily focused on disinformation affecting minority communities; for example, it worked with the Asian American Disinformation Roundtable to produce a report on narratives affecting Asian

Americans during the COVID-19 pandemic.²¹ That topic presents challenges to many researchers who are less familiar with diaspora dynamics, languages, and even the different social media and messaging applications used by those communities. This is the kind of gap DDL’s members can fill.

20 Disinfo Defense League. (n.d.). [Disinfo Defense League](https://perma.cc/R2L2-GTQ5). [perma.cc/R2L2-GTQ5]

21 Asian American Disinformation Table. (2022, August). [Power, Platforms, Politics: Asian Americans and Disinformation Landscape Report](https://perma.cc/MEQ4-U3VG). [perma.cc/MEQ4-U3VG]

DDL members communicate over a shared listserv, and the League provides training for its member organizations to help them “identify, analyze, and respond to disinformation targeting their communities.” It also coordinates counter-messaging strategies and campaigns, informed by the research and monitoring of DDL’s small team and its members. In December 2021, DDL released a policy platform calling for reforms to the collection and use of personal data by tech companies, the prevention of algorithmic discrimination, greater transparency from social media platforms, protection for whistleblowers from within tech companies, and greater federal oversight and consumer protection of the industry.²²

Other U.S. Examples from Counter-Extremism

Other organizations also monitor online threats to U.S. elections. One perspective not well-captured in the previous examples is counter-extremism research. The Institute for Strategic Dialogue and the Anti-Defamation League (ADL) Center on Extremism both specialize in tracking, monitoring, and analyzing online threats from extremist movements and organizations. For instance, in the 2022 election, ADL monitored candidates who had an “extremist nexus” and who spread false claims about the election.

The Institute for Strategic Dialogue, on the other hand, works with election officials targeted by online threats to help them understand and respond. This can include public communications about misleading narratives or additional security precautions for election workers.

Both of these organizations deal with what an ADL staffer called matters of “life or death”—they track the actors most likely to contribute to election violence. This can involve monitoring “esoteric” parts of the web that serve as hubs for white supremacists, militia groups, and other extremists. Both also report findings to law enforcement when they reach a certain threshold of risk for offline violence.

22 Disinfo Defense League. (2021, December). [Disinfo Defense League Policy Platform](https://perma.cc/2ZXW-CYMX). [perma.cc/2ZXW-CYMX]

International Examples

Other models for this work come from outside the United States. Like U.S.-based models, they vary in size, scope, and approach.

The European Digital Media Observatory (EDMO)

EDMO is a large network of researchers, fact-checkers, and media literacy experts focused on disinformation narratives, their analysis, and how to build a resilient society. It is a step forward from the Social Observatory for Disinformation and Social Media Analysis, a community of fact-checkers, researchers,

media literacy professionals, and other counter-disinformation initiatives in the European Union.

Today EDMO comprises more than one hundred partner organizations, with fourteen hubs covering every EU member state and Norway (some hubs are multinational). The hubs have monitored national EU elections and are working together to monitor disinformation trends and narratives in the EU 2024 elections.

Today EDMO comprises more than one hundred partner organizations, with fourteen hubs covering every EU member state and Norway.

EDMO's work also informs policymakers, especially around researcher access to platform data. EDMO has been working to design the framework²³ for the establishment of an independent intermediary body that will administer voluntary data sharing by online platforms under the EU's General Data Protection Regulation; this independent intermediary body may also be able to vet researchers who apply to access platform data under the EU Code of Practice on Disinformation and the Digital Services Act—a widely anticipated development that will facilitate academic research on social media and online harms.

23 European Digital Media Observatory. (2022, May 31). [Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access.](https://perma.cc/PZ73-V3ZL) [perma.cc/PZ73-V3ZL]

In an interview, EDMO Secretary General Paula Gori said that all EDMO's findings are publicly available on its website. EDMO interacts with all relevant stakeholders, including with online platforms, through both public forums like academic conferences and through EU mechanisms like the voluntary code of practice on disinformation (for which EDMO joined the permanent task force). Gori said this was a "win-win" arrangement because platforms welcomed clarity on rules and the European Commission engaged them with civil society, fact-checkers, advertisers, and other relevant stakeholders in co-regulation.

The Carter Center

The Carter Center, on the other hand, is a U.S.-based organization known for its election monitoring efforts abroad. As part of these efforts, it monitors the media environment in the lead-up to an election, including social media. The Carter Center typically hires a small team of local contractors to track misleading narratives and campaign rhetoric across the most relevant platforms in a given context for the purpose of reporting on the electoral environment and the fairness of the vote. It does not typically issue recommendations to platforms or policymakers or interact with law enforcement.

Lessons from Brazil

Brazil's disinformation challenges have many parallels to the United States: a polarized, diverse society, populist leaders, and a presidential transition marred by post-election violence. In 2018, the year former Brazilian President Jair Bolsonaro was elected, Brazilian researchers studied a wave of pro-Bolsonaro organizing on WhatsApp. João Guilherme Bastos Dos Santos, a Brazilian scholar, said he monitored ninety WhatsApp groups in 2018, many of which were connected to the Brazilian far-right. WhatsApp reached out to Dos Santos because of his research, and he met with the head of policy for Brazil.

But in the end, he felt his research and recommendations did not gain traction. He suggested, for example, that WhatsApp act to limit the size of groups in order to slow the spread of content, which

often relies on large groups to go viral. But from 2018 onward, he found WhatsApp's competition with Telegram to be a more powerful driver of policy than Brazilian electoral integrity.²⁴

For the 2022 election, Dos Santos said he moved "closer to applied research dealing with threats against democracy." In 2018, he wanted to show platforms what they could do to avoid spreading disinformation; but in 2022, he was much more concerned about applying his findings in order to avoid a potential coup d'état. Concerned that Bolsonaro would not accept the election results, Dos Santos joined a group of twenty researchers monitoring YouTube, Instagram, TikTok, Twitter, Telegram, and other platforms as part of a project called "Democracia em Xequê," or "Democracy in Check."²⁵ They created real-time dashboards showing disinformation narratives and developed tools to study how YouTube recommendation algorithms promoted far-right content even when search results did not.

Dos Santos also determined it would be more productive to work with the Brazilian Superior Electoral Court than with platforms. The Court had made protecting the election from digital disinformation a top priority and issued steep fines and rulings on content moderation. This brought even some of the most recalcitrant platforms to the table: Telegram, for example, initially ignored their outreach completely until a government minister threatened to ban the app. After that, Telegram began removing election disinformation groups "systematically." Dos Santos took note,

-
- 24 In 2018, WhatsApp was similarly asked to reduce the limit on message forwarding from twenty to five but publicly declined to do so. In 2022, it announced plans to raise the maximum number of group members from 256 to 512, but delayed this and other features like large file-sharing in Brazil until after the 2022 election. See: Haynes, B. & Boadle, A. (2018, October 23). [Despite Brazil election turmoil, Facebook stands by WhatsApp limits.](https://www.reuters.com/technology/whatsapp-stands-by-whatsapp-limits-2018-10-23/) *Reuters*. [perma.cc/A2TK-V37G]; Mari, A. (2022, May 6). [WhatsApp to postpone roll-out of larger groups in Brazil.](https://www.zdnet.com/article/whatsapp-to-postpone-roll-out-of-larger-groups-in-brazil/) *ZDNET*. [perma.cc/9XLN-2RQL]
- 25 "Check" is a position in chess where a player's king is threatened by another piece, but can avoid capture; "Checkmate" is a situation where the king cannot avoid capture on the next turn, resulting in the game's end.

concluding that, “if the Electoral Court needed to go that hard,” researchers had a much better chance at impact if they worked with the government to influence platforms.

He has also found, though, that independent researchers provide key expertise to government actors who otherwise propose policies with negative consequences for free expression. “People from the judiciary don’t really know how platforms work,” he said, “and they can take actions with bad side effects.”

Comparing and Contrasting Counter-Election-Disinformation Initiatives

The varying approaches that election integrity initiatives take reflect distinct perspectives on how to protect elections. In interviews with relevant individuals, participants from the three third-party partnerships articulated three main goals that defined their approaches: research, policy advocacy to both platforms and government, and enabling rapid response in various forms such as policy enforcement by platforms, security measures by government, and fact-checking or counter-messaging aimed at increasing public resilience to misleading or harmful content.

Rapid Response is a Shared Goal

The EIP, Common Cause, and DDL all consider improving some form of rapid response to be part of their work. For example, the EIP described one of its objectives as identifying misinformation before it goes viral—but also says that it does so in order to inform election officials and civil society so they can respond with accurate, authoritative information. The DDL bolstered its members’ work by developing and organizing counter-messaging campaigns in response to disinformation; it also organized a tipline where individuals can report it (similar to Common Cause’s reliance on

volunteers to track disinformation narratives) and circulated a “Disinfo Defense Toolkit” for organizers and advocates.²⁶ Common Cause used its network of volunteers in more than thirty states to similar effect, identifying emerging narratives on- and offline in order to equip the public and election officials with situational awareness and potential counter-messages.

Some Groups Focus on Advocacy, Some on Research

While some initiatives orient their research to further their advocacy goals, others focus on research and may even eschew the label of “advocate.” Researchers by and large aim to enable better responses from stakeholders in social media platforms and government by providing additional monitoring capacity and important independent perspectives. At least some members of the EIP view research as its primary purpose and consider the implications of that research for platform or public policy as secondary, if at all.

While Common Cause, like the EIP, monitored and analyzed misleading content and reported it to platform staff, it also openly embraced an advocacy role. Common Cause acts less as a partner to technology companies and more as a force for their accountability, reporting potential terms of service violations and election threats to platforms as a means of identifying gaps in social media policy, pressuring companies to fill those gaps, and calling for them to better enforce policies already on the books. It also uses its findings to call on policymakers to consider legislative and regulatory action.

More traditional election monitors have similarly combined research and advocacy. For example, the Carter Center used Crowdtangle data and NewsGuard rankings of news site reliability to produce an analysis of election disinformation during the 2020 election and in

26 PEN America. (2020, October 27). [The Fight Against Disinformation Requires the Right Tools](https://www.penamercanet.org/press-releases/the-fight-against-disinformation-requires-the-right-tools). [perma.cc/B7RB-QKTZ]

the lead-up to the January 6th insurrection, complete with sixteen recommendations for platforms to improve their integrity efforts.²⁷ In an interview for this report, the Carter Center said it does not analyze data in real-time, but might alert election authorities of especially concerning false claims about an election.

In stark contrast to both the EIP and Common Cause, the DDL does not interface directly with platforms at all, though some of its members do. According to one individual involved, DDL instead seeks to provide “connective tissue” for civil society through capacity-raising and advocacy functions. It also produced education initiatives to protect local communities and provides policy recommendations to lawmakers.

Are Independent Counter-Election-Disinformation Initiatives Providing Free Labor for Multi-Billion Dollar Corporations?

DDL’s decision not to interface with platforms stems from its judgment that those partnerships are unlikely to germinate into what it sees as necessary reforms. In interviews, both DDL and Common Cause expressed reservations about relationships between platforms and civil society. For Jesse Littlewood, Vice President for Campaigns at Common Cause, his reservations come from apprehension about conducting what disinformation scholar Joan Donovan once called “glorified content moderation for companies valued in the billions”—a form of free labor provided to corporations by scholars and advocates with limited grant funding.²⁸ Littlewood called this relationship “inappropriate” over the long

27 Baldassaro, M., Harbath, K., & Scholtens, M. (2021, August). The Big Lie and Big Tech: Misinformation Repeat Offenders and Social Media in the 2020 U.S. Election. *The Carter Center*. [perma.cc/FJ3G-ZJBR]

28 Donovan, J. (2021, January). Shhhh... Combating the Cacophony of Content with Librarians. *Global Insights*. [perma.cc/688X-J55E]

term; both he and former Meta Public Policy Director Katie Harbath agreed that civil society can find itself between what Harbath called “a rock and a hard place” when deciding how much effort to spend engaging platforms on content policy.

On the one hand, when that engagement goes well, it can prevent immediate real-world harm. The scale of content moderation is too large for platforms to do perfectly; Harbath pointed out that civil society provides important context and expertise to platforms.

She gave the example of civil society noting that the frog emoji had become a coded symbol for white supremacists. The situational awareness civil society gleans from monitoring election disinformation also equips it to pursue other important activities like counter-messaging and public education.

The situational awareness civil society gleans from monitoring election disinformation also equips it to pursue other important activities like counter-messaging and public education.

On the other hand, Harbath acknowledged those who feel that this work is “extractive,” and Littlewood said it can be hard to strike a balance between engaging productively and perpetuating a system that needs larger reform. He called disinformation monitoring an “unfortunately necessary way to reduce harm,” but not a strategy for solving the problem, and worried that platform engagement with civil society could become akin to corporate “greenwashing” (i.e., the practice by which corporations engage in superficial activities to give the appearance of environmental responsibility without adjusting their practices in more impactful ways).

In the long run, Littlewood hopes civil society can attain a more equal power dynamic with technology companies by creating pressure for reform through public education and policy advocacy. He was excited that reforms in the EU had shifted the sense of what is possible in the regulatory space and felt more public awareness could help achieve reforms in the United States. “If we’re going to have unpaid moderators,” he asked, “can we do it in a way that sets us up for effective advocacy for solving the problem and not just cleaning up the mess over and over?” He believes that advocates, not academics, should take the lead on election disinformation monitoring because it would better equip them with the evidence base and leverage to negotiate changes—an important way of building movement power.

Initiatives' Relationships with Government and Law Enforcement Vary from Routine to None at All

The rising risk of election violence means that by the nature of the social media content they monitor, many initiatives interact with government and law enforcement agencies as well as platforms. Staff at the Institute for Strategic Dialogue and the Anti-Defamation League, for example, said in interviews that they contact law enforcement when violent rhetoric online looks likely to spill into targeted offline violence. Election workers are a frequent target of this kind of threat. Some professionals in this field, however, are wary of relationships with police and federal law enforcement.

Outside observers have also politicized and targeted these relationships. Their efforts culminated in a July 4th, 2023, injunction by a federal judge in Louisiana in the case of *Missouri v. Biden*, in which plaintiffs allege the government violated First Amendment rights by working with platforms and independent researchers to engage in viewpoint censorship. The injunction—which was narrowed by the 5th U.S. Circuit Court of Appeals and at the time of publication has been temporarily stayed—restricted numerous Biden administration officials, including some in federal law enforcement, from interfacing with platforms regarding many forms of online content.²⁹ The lower court injunction had also applied to interactions with independent researchers, including the EIP.

The injunction did include exceptions for public safety and other serious circumstances, but these are not always clear cut. Even if the injunction does not ultimately stand, the chilling effect alone may be harmful to counter-disinformation partnerships with government.³⁰

- 29 MacCarthy, M. (2023, July 13). [Internet referral programs are in urgent need of reform](#). The Brookings Institution. [perma.cc/9UM5-XKU7]; Hsu, T. & Thompson, S.A. (2023, July 5). [Disinformation Researchers Fret About Fallout From Judge's Order](#). *New York Times*. [perma.cc/KD7S-FTX6]; Quinn, M. (2023, July 14). [Appeals court halts order barring Biden administration communications with social media companies](#). *CBS News*. [perma.cc/86KR-VW6W]
- 30 Pierson, B. & Goudsward, A. (2023, July 6). [Order limiting Biden officials' social media outreach on shaky legal ground, experts say](#). *Reuters*. [perma.cc/4GKA-TG28]

Federal agencies and employees are likely to err on the side of caution, significantly curtailing their engagement with platforms and researchers; in fact, the State Department already canceled a standing meeting to discuss foreign threats to the 2024 election despite an exception for national security included in the lower court's order.³¹

Some of the individuals interviewed for this report were wary of giving too much detail about their contacts with law enforcement, but in general, gave the impression their efforts were less routine and more ad hoc than initiatives like the Election Infrastructure Information Sharing and Analysis Center (EI-ISAC), which was considered a valuable opportunity to share observations about threats to elections, including trends in disinformation and violent online mobilization. While no concrete evidence of undue government influence on content moderation through the EI-ISAC has emerged, some interviewees expressed sympathy for the idea that these relationships should be more transparent to guard against future abuse.

Why Are Partnerships With Independent Initiatives Valuable?

Despite reservations about the relationship between platforms and initiatives monitoring election disinformation, several interview participants (including former Meta official Katie Harbath) said that independent initiatives provide important benefits that in-house platform integrity workers cannot, even when their work looks similar.

31 Menn, J., Oremus, W., Zakrzewski, C., & Nix, N. (2023, July 5). State Dept. cancels Facebook meetings after judge's 'censorship' ruling. *Washington Post*. [[perma.cc/5HUZ-URU7](https://www.washingtonpost.com/technology/2023/07/05/state-dept-cancels-facebook-meetings-after-judge-censorship-ruling/)]

Interestingly, external researchers also saw platforms as providing important quality controls on their own work: a former researcher at the Digital Forensic Research Lab, for instance, recalled an instance where a staffer took a dataset of suspected inauthentic activity to Twitter, which was able to disprove these suspicions.³²

Independent Counter-Election-Disinformation Efforts Make Platforms More Capable and Accountable

There should be no question that relationships with outside experts are valuable for industry. While platforms enjoy immense financial resources, they cannot create expertise out of thin air; it requires time as well as money. Harbath alluded to this when interviewed, recalling the historical context around Meta's relationship with DFRLab when Facebook teams dedicated to election integrity and countering influence operations were just getting off the ground in 2017. DFRLab and groups like it brought outside perspectives and additional capacity which were valuable. Glenn Ellingson at the Integrity Institute, who previously served on Meta's Civic Misinformation team, said that inbound leads from civil society organizations are especially helpful in contexts where Meta lacks competency, such as the Rohingya genocide in Myanmar.

Another former Meta employee who worked on terrorism and violent extremism issues agreed that third-party coalitions can be a central space for vetting judgments and policy decisions. They said third-party experts are a useful external reference for companies who ultimately "don't want responsibility" for these kinds of decisions and would prefer to justify their actions through third-party assessments. This is an important alternative to the other main way in which platforms bring in expertise quickly: hiring contractors. Platforms may prefer this form of outsourcing because

32 DFRLab. (2023, March 3). [The DFRLab responds to "Twitter Files" story](https://perma.cc/7DSN-2CJE). Medium. [perma.cc/7DSN-2CJE]

Third-party groups play an important role in fostering accountability and transparency while maintaining a firewall between the state and free speech; they serve as both a watchdog and a source of consultation.

it is both quicker than building internal capacity and easier to scale up (or down) from a business standpoint. As these contractors work across the industry, they end up serving as information brokers, and their judgment calls about what does, or does not, violate policy can play an important role in setting precedent without a lot of

scrutiny. In the words of this former Meta employee, “If you let [contractors] make the calls for you, you’ve outsourced policy in addition to monitoring.” Partnering with outside experts from academia and civil society can play a similar role, with more accountability to the public interest.

As one researcher said, it is important for there to be oversight of social media content moderation that comes from beyond the industry (which faces incentives against transparency) or from the government (which faces temptations to overreach or abuse its power). Third-party groups play an important role in fostering accountability and transparency while maintaining a firewall between the state and free speech; they serve as both a watchdog and a source of consultation. Several interview participants said that validation from third-party researchers is a useful response to allegations of platform censorship—including former Facebook officials who pointed to Meta CEO Mark Zuckerberg’s statement that social media companies should not be “arbiters of truth.”

Others noted, however, that civil society organizations are ultimately not accountable to the public. Despite this, they play a significant role in deciding whose concerns are elevated to platform staff and election officials. This is an imperfect model for external oversight, and the implication that these organizations lean to the political left only strengthens unfounded claims of censorial conspiracies. Such allegations are already feeding public cynicism: the recent attacks on counter-election-disinformation initiatives seem likely to be motivated in part by the desire to diminish the credibility of independent counter-election-disinformation initiatives and content moderation as a whole.

Independent Initiatives Bring Language and Cultural Competency That Platforms Need

Independent election disinformation researchers can also help platforms fill capacity gaps around language and cultural context. Multiple researchers interviewed for this report noted that content moderation is markedly worse in non-English languages. One said that “nothing is heavily monitored,” but English is the most well-covered language.

In the 2020 election, disinformation circulating in Spanish was of special concern to platforms, advocates, and researchers. Interviewees who monitor Spanish-language social media said that many of the disinformation narratives in Spanish and English are the same; the different experience of the two languages largely results from the unequal quality of content moderation. They said Spanish-language disinformation is largely unchecked: on Facebook, TikTok, and YouTube, they found content that was removed in English but remained online in Spanish. The researchers gave a possible technical reason for this: the keyword searches some platforms rely on are less useful in Spanish than in English because the language tends to be more “phrase-based,” and there are many different ways to say the same thing. The researchers provided memos to platforms helping them navigate this and other challenges, but platforms did not always take action on violating content included in these memos. This was true even in some instances when the same content was removed in English.

External Partners Can Provide Valuable Visibility and Communication Across the Broader Field

Interviews for this report also surfaced two ways in which independent initiatives can share information across sectors and between companies. First, one Meta staffer interviewed for this report remarked that these partners interface with governments,

other civil society organizations, and the public in ways that platforms cannot (or do not), making them a valuable part of the field's communications infrastructure. Others described how external researchers can follow online harms as they "bounce from platform to platform." Bad actors today often use multiple platforms for different purposes: they may coordinate on Telegram, for instance, then popularize narratives on Facebook or Twitter before drawing audiences onto less well-moderated platforms like Parler, where they share election-related content and calls for violence not permissible elsewhere.

While staffers or contractors working for a platform are often primarily concerned with activity affecting that platform, external researchers can look across the broader ecosystem—though platform staff were quick to point out that some companies, like Meta, do have internal teams looking for threats on other platforms. Still, the EIP considered their analysis of inconsistent platform policies to the same threats across platforms an important success and a key means of identifying gaps in online trust and safety.³³

Working in Coalitions Brings Many Benefits, But Also Some Risks

Coalitions of initiatives like the EIP and DDL share an emphasis on coordination and collaboration between individuals and organizations. Researchers say these coalitions give them better access to timely peer review—an important quality control measure. They also noted that, given the use of targeted lawsuits to deter and distract disinformation researchers, coalitions can provide insulation from lawsuits. To give an example, when all the members of the EIP sign their names to a publication, it leaves them less vulnerable to reprisals from politically motivated actors.

33 The Election Integrity Partnership. (2021, June 15). [The Long Fuse: Misinformation and the 2020 Election](https://perma.cc/REV4-6Q9A). [perma.cc/REV4-6Q9A]

Katie Harbath noted that different researchers bring different data and skill sets to the problem, from which all of them and platforms benefit when they work together. From a practical standpoint, she also said it's simply easier to have a smaller number of central contact points—for both platform staff and the researchers trying to contact them.

Another trust & safety professional at Twitter made similar remarks: Working through an umbrella organization helps streamline the otherwise unwieldy process of vetting and signing contracts and nondisclosure agreements with a large number of partners. This is especially true when the coalition involves dozens of organizations: consider that while the EIP involved four institutions with prior relationships with one another and platform staff, the European Digital Media Observatory (previously the Social Observatory for Disinformation and Social Media Analysis) includes more than 120 institutions across EU member states and sectors like media, fact-checking, and academia.³⁴ The diversity of EDMO's membership means that the relationship is greatly streamlined by the presence of one body that can represent many stakeholders.

However, large, diverse coalitions introduce an element of risk: a dramatic shift in the professionalism or political orientation of one member can harm the credibility of the broader coalition and its relationships with other stakeholders.

34 European Digital Media Observatory. (n.d.). [List of Institutions connected to EDMO.](https://perma.cc/35ML-7E27) [perma.cc/35ML-7E27]

03

Challenges for Independent Counter-Election- Disinformation Initiatives

Despite the innovation and benefits demonstrated by these initiatives in all their variety, today they face major obstacles that threaten their effectiveness. Frustrations between researchers and social media companies have only grown more serious since the 2020 and 2022 elections and the layoffs across the tech industry. Platforms have changed policies in ways that make combatting election mis- and disinformation more difficult. New platforms complicate the task of monitoring the internet's bad actors and the political climate for disinformation research has become increasingly stormy as election denial becomes central to Republican politics and the party's efforts to reclaim the White House in 2024.

Partnerships and Approaches Are Too Time-Bounded

As one researcher at Common Cause said, “There is no time when election disinformation isn’t impacting access to the vote.”

Platform integrity teams have been criticized for opacity around when special efforts to protect elections start, and—crucially given violence weeks after the elections in the United States and Brazil—when those efforts end.³⁵ Some observers say these efforts need to be made permanent rather than dialed up and down around election periods. Terms of service are also typically more strict about false claims regarding the time and manner of voting than they are about denial of an election’s legitimacy or outcome. When platforms do have policies against election delegitimization, they are usually time-bound, limiting the value of engagement between civil society and platforms at times when voting is not ongoing.

Similarly, our interviews found consensus that independent efforts to counter election disinformation should be ongoing. Interview participants who worked at platforms said that sustained engagement promotes long-term, trusting relationships. One said that persistent engagement might also alleviate concerns that “election-timed organizations” are necessarily political even when they strive for non-partisanship.

But independent researchers gave the most compelling argument: their work needs to be persistent because election disinformation itself is persistent, though its frequency and pitch rise and fall. As one researcher at Common Cause said, “There is no time when election disinformation isn’t impacting access to the vote.” While Common Cause “scales up” its efforts during election season, since

35 Bhatia, A. & Adler, W.T. (2023, March 23). [CDT Weighs In on Meta Oversight Board’s case on Takedown of Speech Calling for Attack on Brazil’s National Congress](https://perma.cc/83GC-TP9Y). *Center for Democracy & Technology*. [perma.cc/83GC-TP9Y]

2020 it has engaged more consistently with its partners. Not only do primaries and local elections occur frequently in the election off-season, but lawmakers also propose legislation that could restrict voting access.³⁶

Jesse Littlewood from Common Cause said election disinformation lays the narrative groundwork for these legislative and campaign pushes, creating a predicate for new restrictions on voting and attacks on nonpartisan election administration. According to the Brennan Center for Justice, “more than 440 bills with provisions that restrict voting access” were introduced in 2021.³⁷ The pace of this activity is increasing: The Brennan Center writes that 150 such bills were introduced across at least 32 states in the first two months of 2023 alone—more than the same point in either 2021 or 2022.³⁸ In October 2022, the *New York Times* reported that more than 370 candidates for office in the midterm elections that year cast doubt on the legitimacy of the U.S. 2020 election.³⁹ While most of the highest-profile election deniers ultimately lost their races, at least 177 did not.⁴⁰

Littlewood said capacity constraints are another reason organizations like Common Cause do not typically sustain the same tempo of activity during elections as they do between them. Funding for this work becomes more available in the run-up to national elections, and volunteers become more engaged during the same period. Capacity issues also affect similar efforts; the EIP

36 This has primarily been true at the state level, but proposed legislation in the House of Representatives would federalize the trend: Brower, M. (2023, July 10). [House Republicans Unveil Most Restrictive Elections Bill in Decades](#). *Democracy Docket*. [perma.cc/TT3F-9G8D]

37 Brennan Center for Justice. (2022, January 12). [Voting Laws Roundup: December 2021](#). [perma.cc/AAV9-PGUK]

38 Brennan Center for Justice. (2023, February 27). [Voting Laws Roundup: February 2023](#). [perma.cc/CN3C-8ERN]

39 Yourish, K., Ivory, D., Byrd, A., Cai, W., Corasaniti, N., Felling, M., Taylor, R., & Weisman, J. (2022, October 13). [Over 370 Republican Candidates Have Cast Doubt on the 2020 Election](#). *New York Times*. [perma.cc/8J42-PGQU]

40 Gallagher, K. (2022, December 13). [The most prominent election-deniers lost their races. But at least 177 have won so far](#). *Insider*. [perma.cc/9APB-84NY]

for instance was an intensive round-the-clock monitoring effort that would require a significant bump in resources to sustain year round. The DDL said that the core mission of many of its members is not countering disinformation, and so outside of election season many member organizations focus on other issues relevant to their constituencies.

Platforms Are Becoming Less Responsive

Interview participants complained that platforms have become less communicative since the 2020 election, especially since widespread layoffs in the tech sector beginning in 2022. Even before the layoffs, however, research and reporting found widening inconsistencies

in how individual platforms enforced policies around election misinformation between 2020 and 2022.⁴¹

What's more, since 2022 platforms have loosened policies against election misinformation, sometimes in ways that essentially capitulate on the issue.⁴²

Delays in detection and moderation mean that even when independent efforts succeed in motivating platforms to act, they are often too late to contain much of the damage.

These problems are especially severe considering that former Meta engineers say that, when content goes viral, the surge in views typically comes early. Therefore, speed is of the essence. Delays in detection and moderation mean that even when independent efforts succeed in motivating platforms to act, they are often too late to contain much of the damage. This makes the declining responsiveness of platforms to outside researchers all the more concerning.

- 41 Bradshaw, S. & Grossman, S. (2022, August 7). Were Facebook and Twitter Consistent in Labeling Misleading Posts During the 2020 Election? *Lawfare*. [perma.cc/9MXZ-8ESE]; Nix, N., Merrill, J.B., & Godfrey, H. (2022, November 6). This year, GOP election deniers got a free pass from Twitter and Facebook. *Washington Post*. [perma.cc/8WJ2-P6SS]
- 42 Ingram, M. (2023, June 15). The tech platforms have surrendered in the fight over election-related misinformation. *Columbia Journalism Review*. [perma.cc/U8F9-E4RX]

Staff at Common Cause complained that when they found instances of viral disinformation, they reported it to contacts at those companies—but too often, platform staff did not follow through on these reports. Other researchers had similar experiences, and some said they now only report content to platforms in the most extreme circumstances when it obviously violates policy. They cited the example of attorney and conspiracy theorist Lin Wood, who they reported to Twitter multiple times but whose account was not suspended until after the January 6th attack on the U.S. Capitol.⁴³

A September 2021 report from Common Cause claims that platform responsiveness deteriorated further in the months following the insurrection as election misinformation that would have previously been removed or labeled by platforms went unmoderated.⁴⁴

In an interview, a Common Cause researcher shared an example that illustrates how this trend continued through the 2022 midterms. Common Cause reported the spread on Twitter, Facebook, and TikTok of “wanted posters” featuring images of election workers accusing them of facilitating election fraud and encouraging viewers to identify and report them to law enforcement. The companies did not act on the report, so Common Cause provided it to ProPublica, which published a story about it.⁴⁵ Afterward, the companies took action on the posts, and three months later, Meta updated its policies on harassment to clearly include “directed mass harassment” against election officials.⁴⁶

- 43 Wood was initially temporarily suspended for violent incitement, but was banned permanently after he said publicly he would evade Twitter’s content moderation by creating a second account. That Wood was only given a temporary suspension after January 6th even though his tweets were flagged for Twitter in the weeks before illustrates the degree of inconsistency in content moderation generally. Mac, R. (2021, January 7). [Trump-Supporting Lawyer Lin Wood Has Been Permanently Banned From Twitter](https://www.buzzfeednews.com/article/rmac/trump-supporting-lawyer-lin-wood-has-been-permanently-banned-from-twitter). *BuzzFeed News*. [perma.cc/AX76-6CCA]
- 44 Littlewood, J., & Steiner, E. (2021, September 2). [Trending in the Wrong Direction: Social Media Platforms’ Declining Enforcement of Voting Disinformation](https://www.commoncauseeducation.org/2021/09/02/trending-in-the-wrong-direction-social-media-platforms-declining-enforcement-of-voting-disinformation/). *Common Cause Education Fund*. [perma.cc/3QDT-YAHT]
- 45 Kroll, A. (2022, June 17). [“Big Lie” Vigilantism Is on the Rise. Big Tech Is Failing to Respond](https://www.propublica.com/big-lie-vigilantism-is-on-the-rise-big-tech-is-failing-to-respond/). *ProPublica*. [perma.cc/95JV-T95C]
- 46 Meta. (n.d.). [Facebook Community Standards: Bullying and Harassment](https://www.facebook.com/communitystandards/bullying-and-harassment/). [perma.cc/G64M-ZCWV]



Figure 8. Image reported by Common Cause to platforms and eventually published by ProPublica after platforms did not act on the report.

ProPublica captured this screenshot and added a red label to avoid incidentally misleading readers who view the image.

Source: Kroll, A. (2022, June 17). *"Big Lie" Vigilantism Is on the Rise. Big Tech Is Failing to Respond.* ProPublica. [perma.cc/95JV-T95C]





Figure 9. Screenshot of Facebook terms of service.

The green highlighting is part of Facebook's change log and shows that "election officials" were added to a list of individuals at heightened risk of offline harm on September 29, 2022.

Source: Meta. (n.d.). *Facebook Community Standards: Bullying and Harassment*. [perma.cc/G64M-ZCWV]

- Remove directed mass harassment, when:
 - Targeting, via any surface, 'individuals at heightened risk of offline harm', defined as:
 - Human rights defenders
 - Minors
 - Victims of violent events/tragedies
 - Opposition figures in at-risk countries during election periods
 - Election officials
 - Government dissidents who have been targeted based on their dissident status
 - Ethnic and religious minorities in conflict zones
 - Member of a designated and recognizable at-risk group

The ProPublica example demonstrates a marked lack of trust between platforms and many outside researchers, especially around relationships with the press. A researcher at Common Cause called this "one of [their] only social media wins. "For whatever reason," they said, "we have to be the ones to point it out." They described Twitter as similarly "reactive" and said they believed change at platforms too often required someone to "make a fuss" in the press. Jesse Littlewood from Common Cause said that in his experience groups do not use media attention except in specific situations where platforms don't take action quickly, and described the decision to do so in the ProPublica example as a rare "break-glass" measure taken because of the risk of imminent political violence.

When asked about examples like these, Katie Harbath provided several points from a platform perspective. First, she said, platforms do not have unlimited capacity for content moderation, which cannot be purchased off the shelf but requires time to train personnel, so they are constantly required to triage. Sometimes reported content does not reach the front of the line. Sometimes this is because platforms are focused on other problems that are not visible to external observers; sometimes, it is because opinions differ between platform teams about whether or not a policy was violated. Opinions also differ between external researchers and platforms—for instance, in the example above, it is possible Facebook staff did not see the wanted posters as violating policy because election officials were not defined as "at heightened risk." Following the letter of the existing policy, they may have allowed the post to remain up despite what Common Cause saw as a

clear case of targeted abuse. Harbath said that press attention can bring this type of content to the attention of higher-level decision-makers, who can make difficult judgment calls about edge cases or necessary policy changes.

Many of the researchers interviewed for this report also noted the spottiness of their contact with platforms. For instance, when Meta approached a large research nonprofit with a proposed three-tier escalation system for threats of race-based hate violence, the researchers worked with the platform to identify potential gaps and weaknesses. After that meeting, Meta never circled back and the researchers never learned if those insights were put to use. Platform staff interviewed for this report said that sometimes, this happens because platform lawyers oppose further engagement on a topic—suggesting the issue is not just continuity of contact but also who is influential within the company itself.

While these types of complaints have been common for years, there are reasons to believe the problem grew worse as the 2020 election receded into memory. In October 2021, Facebook rebranded itself as Meta in a show of commitment to its vision of a “metaverse” using new technology to connect workers and consumers across digital spaces. Since then, reports from inside the company indicated that CEO Mark Zuckerberg stopped considering election integrity a top priority and stopped meeting with the elections team.⁴⁷ Glenn Ellingson, a former Meta integrity worker, pushed back slightly on the notion that Zuckerberg’s disengagement is a sign of a broader problem—in his view, being the primary focus of the chief executive is more disruptive than it is helpful. But the experience of independent researchers suggests the broader pullback is real.

Ellingson also said that external researchers often had specific points of contact at Meta whom they could reach out to with research leads. However, Katie Harbath noted that one of the biggest challenges for external researchers when interacting with platforms is knowing which team to contact with relevant queries. The organizational chart of many social media companies is opaque to outsiders, and Meta’s in particular is known to be

47 Frenkel, S., & Kang, C. (2022, June 23). [As Midterms Loom, Elections Are No Longer Top Priority for Meta C.E.O.](https://www.nytimes.com/2022/06/23/us/politics/meta-election-integrity.html) *New York Times*. [perma.cc/YF7T-XBHJ]

sprawling. Another former Meta staffer said that it often falls to the team that receives leads from partners to route those leads to the correct team—which may or may not find a policy violation within the provided data. This staffer wished that Meta provided more staffing and resources to help triage incoming leads and work with partners to reduce friction—especially since the personnel routing those requests are often “underwater with other work” and spend considerable time verifying incoming leads.

Jesse Littlewood at Common Cause said that, despite these pain points, Common Cause’s relationships with Meta were unusually institutionalized with more consistent points of contact. With other platforms, problems were more likely to arise because relationships were highly individual: if a point of contact left a company, a new relationship would have to begin from scratch. Sometimes the best point of contact is an “info at” email inbox. Littlewood said in most cases it is up to civil society organizations to maintain and replace relationships within platforms—despite the enormous resource gap between organizations in the corporate and nonprofit sectors.

The situation at Twitter is especially severe: Observers say the company’s civic integrity policies have significantly eroded since Elon Musk’s purchase closed.⁴⁸ In an interview, EDMO General Secretary Paula Gori described having contacts come and go at Twitter constantly as personnel changes roiled the company—depriving EDMO of consistent contacts with which to discuss complex EU regulatory issues. Twitter also left the voluntary EU Code of Practice on Disinformation,⁴⁹ a co- or self-regulatory framework consisting of specific actions intended to reduce the spread of disinformation online, limiting its application across the social media industry.⁵⁰

48 Paul, K., & Dang, S. (2022, November 8). [Elon Musk’s Twitter slow to act on misleading U.S. election content, experts say](https://www.reuters.com/technology/twitter-slow-to-act-on-misleading-u-s-election-content-experts-say-2022-11-08/). *Reuters*. [perma.cc/Q8MH-ZZXD]

49 European Commission. (n.d.). [The 2022 Code of Practice on Disinformation](https://ec.europa.eu/commission/presscorner/detail/en/ip22_1747). [perma.cc/Q93E-KJEE]

50 Pitchers, C. (2023, May 6). [Twitter has chosen ‘confrontation’ with Brussels over disinformation code of conduct](https://www.euronews.com/en/2023/05/06/twitter-has-chosen-confrontation-with-brussels-over-disinformation-code-of-conduct). *Euronews*. [perma.cc/Q97S-VVE5]; Twitter’s last transparency report under the Code of Practice, filed in January 2023, was perfunctory—consisting mostly of publicly available information about Twitter policies and claims that most commitments under the code are not relevant to its “current approach.”; Transparency Centre. (2023, January). [Reports Archive](https://www.transparencycentre.org/reports). [perma.cc/N4MT-HWBK]

Tech Layoffs Exacerbate Challenges for Independent Counter-Election-Disinformation Initiatives

Whatever frustrations existed within partnerships between platforms and external research coalitions during the 2020 election, almost everyone interviewed agreed that widespread layoffs in the tech sector in 2022 made the challenges more severe in both degree and kind. These problems have escalated to the point that some researchers are engaged in a wholesale rethink of their approach.

The situation at Twitter is especially severe: Observers say the company's civic integrity policies have significantly eroded since Elon Musk's purchase closed.

In 2022, over one hundred and seventy-five thousand employees working in tech experienced job cuts.⁵¹ In the first two months of 2023, companies cut over 100,000 more.⁵² But the problem is especially acute at Twitter. Just over a week after finalizing his acquisition, Musk laid off 3,700 employees—half of Twitter's entire staff. Fifteen percent of the Trust & Safety team was cut, and other teams—including one focused on human rights and global conflict—were eliminated entirely.⁵³ Coming days before the 2022 U.S. midterm elections, critics alleged the layoffs were reckless and haphazard. They had immediate consequences as staff were locked out of content moderation tools, external partners were unable to contact relevant personnel, and accounts began testing the waters of "new Twitter." Within twelve hours of Musk's Twitter takeover, racist hate speech surged.⁵⁴ In the months since, hundreds

51 Lee, R. (n.d.). [Layoffs.fyi - Tech Layoff Tracker and Startup Layoff Lists](#). [perma.cc/NLC3-CNT4]

52 McCorvey, J.J. (2023, February 10). [Tech layoffs shrink 'trust and safety' teams, raising fears of backsliding efforts to curb online abuse](#). *NBC News*. [perma.cc/R9MW-PLQR]

53 Harwell, D., Zakrzewski, C., & Stanley-Becker, I. (2022, November 4). [Twitter layoffs gutted election information teams days before midterms](#). *Washington Post*. [perma.cc/SN2E-ZHJ8]; Ortutay, B., & O'Brien, M. (2022, November 5). [Elon Musk defends Twitter layoffs as critics see a 'lack of care and thoughtfulness'](#). *Fortune*. [perma.cc/Q4K3-RFU4]

54 Frenkel, S., & Conger, K. (2022, December 2). [Hate Speech's Rise on Twitter Is Unprecedented, Researchers Find](#). *New York Times*. [perma.cc/C4FG-YTPR]

more Twitter employees have quit or been let go, and the company dissolved its Trust & Safety Council, which provided advice to Twitter on reducing online harms.⁵⁵

In an interview for this report, a trust and safety staffer who remains at Twitter reflected on the ramifications of these developments. They said that the past several years have seen an “arms race” between platforms and bad actors as companies responded to Russian interference in the 2016 U.S. elections and scaled those defenses globally. Adjusting to these heightened defenses, disinformation campaigns underwent “incredible iterative evolution,” for example by making their work harder to attribute through outsourcing and by using AI-generated profile pictures to make fake accounts less obvious.⁵⁶ These adaptations marked an era of competition between investigators and bad actors as each worked to get ahead of the other. Now, the Twitter staffer said, “that era has ended”; the new era will be marked by “retrenchment or even retreat.”

While the situation at Twitter is markedly dire, this staffer believes other companies are facing a similar situation. At Alphabet, for example, the *New York Times* reported in early 2023 that cuts left only a single person in charge of misinformation policy for the entire world—reflecting what external researchers called a pattern of disengagement across the industry.⁵⁷

The Twitter staffer said that trust and safety teams across the industry now have gaps in expertise—staff who were once specialists in specific harms must now act as generalists, degrading platforms’ technical, geographic, linguistic, and other knowledge. This diminishment means platforms have lost early warning

- 55 Sixteen former members of the Twitter Trust and Safety Council. (2022, December 14). Joint Statement on the Disbanding of the Twitter Trust and Safety Council. *Center for Democracy & Technology*. [perma.cc/B2YQ-TG69]
- 56 Goldstein, J.A. & Grossman, S. (2021, January 4). How disinformation evolved in 2020. *Brookings Institution*. [perma.cc/D82B-LJ52]
- 57 Myers, S.L. & Grant, N. (2023, February 14). Combating Disinformation Wanes at Social Media Giants. *New York Times*. [perma.cc/3WU9-HJMK]

capabilities that took years to build. Exacerbating the problem is what they called a “loss of interface”: having multiple teams working on trust and safety issues created a “porous boundary” across which civil society and vulnerable communities could find multiple touchpoints. In a diminished field, those touchpoints are gone. Glenn Ellingson at the Integrity Institute made similar remarks, in particular about Meta’s consolidation of what were previously “highly empowered small teams” into a smaller, more centralized staff.

Not every integrity worker is so pessimistic, though all admit that the field faces difficulty. One Meta staffer noted that their team had largely weathered staff cuts by pointing to new regulations—for example, the Digital Services Act in the European Union—and using

Integrity teams both provide transparency reports required by the Act and respond to the harms it seeks to reduce. In their words, “the threat of regulation justifies the work.”

them to justify their budgets. Integrity teams both provide transparency reports required by the Act and respond to the harms it seeks to reduce. In their words, “the threat of regulation justifies the work.”

Even so, external researchers involved in election integrity coalitions are deeply concerned. One said that every point of contact they had at major companies has left as entire teams were disbanded, making it impossible to sustain effective communication.

This presents a major dilemma for previous approaches to promoting election integrity and corporate accountability. In the words of one advocacy executive, “What do you do if a platform is not responding?” A Common Cause researcher said that while the model of third-party researchers holding platforms accountable by finding and reporting election disinformation wasn’t really working before the layoffs, “now there is not even anyone to contact... clearly something was broken before,” but things are “even worse now.”

Lack of External Access to Platform Data Causes Problems for Both Platforms and Researchers

In interviews for this report, and in previous CDT research,⁵⁸ both platform staff and external researchers have raised frustrations stemming from another challenge: researchers' lack of visibility into platforms. The ability of external researchers to access data about activity on social media varies widely by platform: Twitter, for example, is the subject of many academic studies simply because until recently, it offered open access to its API, allowing researchers easy access to broad amounts of data. Meta, on the other hand, restricts this kind of access and instead offers more limited windows through tools like Crowdtangle and its ad library.

A researcher at Common Cause said in an interview that lack of platform data makes their policy proposals “stabs in the dark.” In other words, the empirical basis for assessing problems and recommending solutions is weaker than it should or could be. A researcher affiliated with the DDL said that the lack of data puts tech policy advocates on uneven ground; they recalled the example of New York Times journalist Kevin Roose, who used Crowdtangle to generate a regular list of the top ten best-performing links on Facebook and analyze their political bent. When he noted that almost every list showed incendiary right-wing commentators doing especially well, Facebook issued a rebuttal saying that Roose’s read on the data was inaccurate—that the top ten most engaged with links are different from and less important than the top ten most

58 Vogus, C. (2022, August 16). [Improving Researcher Access to Digital Data: A Workshop Report](#). *Center for Democracy & Technology*. [perma.cc/QTV2-SA48]

viewed links, data which Roose could not access. Without greater visibility into the platform, researchers struggle to assess claims like these and are more vulnerable to unverifiable contradictions from corporate spokespeople.⁵⁹

From inside the platforms, current and former staffers raised frustrations of their own caused by external researchers' lack of access. Two former Meta employees said that external researchers sometimes seize on less crucial but more visible problems; when this research attracted press attention, they said it could distract platform staff from more dangerous threats less apparent to outsiders. When asked to respond, researchers were quick to say this is a reason to provide data to external researchers, not grounds to criticize their work.

One of the two employees also said that platforms were reluctant to provide data after the Cambridge Analytica scandal—though they acknowledged that data-sharing arrangements could be better structured to prevent abuse. They also worried that if only one platform allowed data access to outside researchers, that platform would be scrutinized far more than its competitors, fairly or unfairly.

The other Meta employee raised examples of mistaken attribution of inauthentic activity by outside researchers; they said researchers should clearly articulate their level of confidence in their findings and not “claim they need special access” to do better work. However, when asked about this in interviews, researchers replied that restricting their work to claims they could make with high confidence given only limited data allows companies to constrain the research agenda, limiting its potential for holding corporations accountable. These examples speak to how the inscrutability of many platforms contributes to growing frustration and lack of trust in relationships between external researchers and platform staff.

59 An important caveat is that not all questions related to online harms require back-end access to platform data to study, and qualitative research is still valuable. But while some researchers call for more mixed methods approaches, hard numbers are powerful for advocacy purposes.

Unfortunately, despite persistent calls for greater transparency into platform data, trends point to less access for most researchers, not more. In 2022, Facebook stopped engineering support for Crowdtangle, beginning a slow process of shutting it down; researchers have since complained the tool has become buggy and the data unreliable.⁶⁰ The Integrity Institute's Glenn Ellingson called the decision not to share data more broadly with researchers "one of the least responsible" decisions made by Meta and other platforms. He believes that governments should not have allowed Meta to purchase and bury Crowdtangle.

In 2021, Meta also suspended the accounts and pages of three NYU researchers who scraped Facebook data to create an "ad observatory" for studying paid content on the platform.⁶¹ In March 2023, Twitter increased the price of access to its API to \$42,000 a month—a rate simply unaffordable for most researchers.⁶² The European Union's Digital Services Act, which will lay out a process through which platforms are obligated to provide vetted researchers with requested data, is a bright spot on an otherwise darkening horizon.⁶³ Similar approaches have been introduced in legislative chambers in the U.S. and Brazil but have not yet passed.

- 60 Lawler, R. (2022, June 23). Meta reportedly plans to shut down CrowdTangle, its tool that tracks popular social media posts. *The Verge*. [perma.cc/QK3G-DKHX]
- 61 The ad observatory is an independent resource distinct from Facebook's ad library; Hatmaker, T. (2021, August 4). Facebook cuts off NYU researcher access, prompting rebuke from lawmakers. *TechCrunch*. [perma.cc/4FWG-TJ7N]; Clark, M. (2021, August 3). Research Cannot Be the Justification for Compromising People's Privacy. *Meta*. [perma.cc/5JPX-X5MM]
- 62 Stokel-Walker, C. (2023, March 10). Twitter's \$42,000-per-Month API Prices Out Nearly Everyone. *Wired*. [perma.cc/2YV5-FR8G]
- 63 For more analysis on the DSA and its data sharing provisions, see: Allen, A. & Stockhem, O. (2022, August 10). A Series on the EU Digital Services Act: Due Diligence in Content Moderation. *Center for Democracy & Technology*. [perma.cc/D5FF-U8GV]; Vogus, C. (2023, January 25). Defending Data: Privacy Protection, Independent Researchers, and Access to Social Media Data in the U.S. and EU. *Center for Democracy & Technology*. [perma.cc/A46F-99HZ]

Researchers have some tools to get around these obstacles, but largely consider them inadequate fixes. DDL, for example, has an “extensive network” of member organizations with visibility into chat threads on apps like WhatsApp and WeChat, giving them better visibility into private messaging spaces (especially those used by minority communities, whose members usually have more cultural and linguistic capacity for analyzing than most researchers).⁶⁴

In international contexts, the Carter Center has used small groups of paid contractors to monitor social media—an approach that also adds knowledge of the local context. An analyst affiliated with the Carter Center said this approach was more suitable for long-term analysis than “real-time” response, though the Center might alert election authorities if they find concerning content. Crucially, the Center also monitors television, radio, and other mediums like podcasts, giving it visibility into larger narratives and dynamics beyond the internet. It has also explored new tools, like Junkipedia, which provides information from fourteen platforms—including TikTok, YouTube, Truth Social, and AM radio.

64 Private messaging applications are notoriously difficult for disinformation researchers to study, especially since they are often encrypted. There are also ethical disagreements between researchers about infiltrating private messaging spaces used to share or coordinate disinformation. It was not clear from the interview how DDL members approach these questions. Some initiatives, like the EIP, explicitly avoid private messaging spaces. Common Cause said in an interview that its volunteers are not asked to enter these spaces because of the potential risks of doing so when monitoring extremist actors. For more information, consider: Goodwin, C. & Jackson, D. (2022, February 9). Global Perspectives on Influence Operations Investigations: Shared Challenges, Unequal Resources. *Carnegie Endowment for International Peace*. [perma.cc/89GG-G88Z]; Kamara, S., Knodel, M., Llansó, E., Nojeim, G., Qin, L., Thakur, D., & Vogus, C. (2021, August 12). Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems. *Center for Democracy & Technology*. [perma.cc/E2LF-S66W]

New, “Alternative” Platforms Complicate Trust & Safety

Changes to the social media landscape are complicating election integrity initiatives in other ways. As the number of noteworthy social media applications rises, researchers must monitor more surfaces and the movement of content and users between them. While early election disinformation monitoring efforts mostly contended with text and image-based content, today the rise of TikTok, the emergence of Instagram stories, and the continued influence of YouTube and Twitch mean that the most popular content is increasingly video-based; as ISD noted in advance of the 2022 midterms, video is more difficult and more time consuming to parse than text.⁶⁵ The pivot to short-form video on platforms like Instagram has only made counter-election-disinformation more difficult.

Katie Harbath said that newer platforms can also struggle to respond to disinformation because they may just be starting to think through trust and safety questions that older platforms have already grappled with. She also said that while attention continues to focus on the largest platforms, it doesn't take a large user base to have a large impact if those users are motivated by extreme views. For example, many of the platforms where users discussed storming the U.S. Capitol in advance of the January 6th insurrection were smaller “alt-tech” platforms. These are also ideologically hostile to content moderation, foreclosing counter-disinformation partnerships with them. Similarly, bad actors increasingly use encrypted private messaging chat rooms to communicate—spaces that can be difficult or even dangerous to infiltrate and monitor.

65 Martiny, C., Jones, I., & Cooper, L. (2022, November 4). [Election disinformation thrives following social media platforms' shift to short-form video content](https://perma.cc/TKM8-PG4A). *Institute for Strategic Dialogue*. [perma.cc/TKM8-PG4A]

As a researcher at the Anti-Defamation League said in an interview, mainstream platforms are still important main hubs for the internet; this is why so many banned accounts returned to Twitter after Elon Musk reversed their suspensions. But the emergence of these new spaces means banning users and certain types of content can have complicated, sometimes counterproductive effects: research following the January 6th insurrection shows that users who are unable to access mainstream platforms often migrate to those where more extreme views are common, which only serves to increase their exposure to extremist content.⁶⁶

Generative AI Brings New Risks

Since the debut of ChatGPT in 2022, generative artificial intelligence—machine learning algorithms that can create unique content from user-provided prompts—has produced a steady stream of press coverage about its implications for trust and safety online.⁶⁷ Some of the researchers interviewed for this report raised concerns about AI-generated disinformation. The use of generative AI in politics is already a reality: In April 2023, the Republican National Committee released a thirty-two-second ad, “Beat Biden,” featuring images of chaos and conflict produced entirely by AI.⁶⁸ That June, Florida Governor Ron DeSantis’s presidential campaign released an ad incorporating AI-produced images of former President Donald Trump embracing former National Institute of Allergy and Infectious Diseases Director Anthony Fauci.⁶⁹ If Americans are so inclined, they can also (as of this writing) watch a never-ending debate between synthetic reproductions of Presidents

66 Buntain, C., Innes, M., Mitts, T., & Shapiro, J. (2023, March 12). [Cross-Platform Reactions to the Post-January 6 Deplatforming](https://perma.cc/8QJA-AW73). *Journal of Quantitative Description: Digital Media*. [perma.cc/8QJA-AW73]

67 Hsu, T., & Thompson, S.A. (2023, February 8). [Disinformation Researchers Raise Alarms About A.I. Chatbots](https://perma.cc/8ZRY-34WZ). *New York Times*. [perma.cc/8ZRY-34WZ]

68 Thompson, A. (2023, April 25). [First look: RNC slams Biden in AI-generated ad](https://perma.cc/5EZ4-4VEE). *Axios*. [perma.cc/5EZ4-4VEE]

69 Sarlin, B. & Talcott, S. (2023, June 8). [DeSantis campaign shares fake Trump/Fauci images, prompting new AI fears](https://perma.cc/F5SZ-SXFE). *Semafor*. [perma.cc/F5SZ-SXFE]

Trump and Biden debating one another using AI-generated speech riffing off comments in the user chat.⁷⁰ (Fair warning: the content is decidedly not safe for work.)

There are at least three arguments that generative AI will supercharge disinformation. The first has to do with cost reduction: the time and resources needed to produce posts about U.S. elections in flawless English fall to near zero with generative AI. Measured by word count, this means the productivity of the average bot or troll could increase dramatically. Second, content produced by AI could be more targeted and persuasive to users, causing it to be shared more often or to mislead more people. And third, sheer volume might make AI-generated disinformation different from what came before. Users who seek to verify an online rumor might be overwhelmed by a vast supply of fake news outlets providing AI-produced falsehoods faster than real journalists can report—problematizing verification strategies in emergency situations.

Each of these three possibilities has counterpoints which, while raised less often, should caution researchers and policymakers from jumping to conclusions about generative AI's future impact on election disinformation. For instance, less costly production of content does not help with its distribution, which some observers note is the main bottleneck for disinformation.⁷¹ Claims that generative AI will be more persuasive echo the early discourse around psychometric ad targeting—now largely considered overhyped.⁷² And, in a crisis, savvy news consumers are likely

70 Robertson, D. (2023, June 21). 'Biden' vs. 'Trump' and the future of debate. *Politico*. [perma.cc/5KE7-QCKF]

71 Kapoor, S. & Narayanan, A. (2023, June 16). How to Prepare for the Deluge of Generative AI on Social Media. *Knight First Amendment Institute at Columbia University*. [perma.cc/E75Z-TZMP]

72 Resnick, B. (2018, March 26). Cambridge Analytica's "psychographic microtargeting": what's bullshit and what's legit. *Vox*. [perma.cc/CC5Z-QVEF]; Krotzek, L.J. (2019). Inside the Voter's Mind: The Effect of Psychometric Microtargeting on Feelings Toward and Propensity to Vote for a Candidate. *International Journal of Communication*. [perma.cc/PLU8-BWE8]; Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. (2022, November 8). Effects of an issue-based microtargeting campaign: A small-scale field experiment in a multi-party setting. *The Information Society*. [perma.cc/7UZZ-X9KX]; Bodó, B., Helberger, N., & de Vreese, C. (2017). Political micro-targeting: a Manchurian candidate or just a dark horse? *Internet Policy Review*. [perma.cc/S9EE-FT93]

to remain skeptical of previously unknown sources and wait for verified reporting to emerge; other users may be in much the same situation as they were before.

It is too early to say with certainty if generative AI will supercharge the age of disinformation in commonly predicted ways—but it will almost certainly have important ramifications for trust and safety.

Concerns about generative AI are reminiscent of past conversations about synthetic video “deepfakes,” which have not (yet) had political consequences at the predicted scale.⁷³ And, as with deepfake video, concerns around generative AI risk wrongfully emphasizing novel technological developments over psychological, social, and political factors that contribute to disinformation’s potency. For instance, a poorly edited “cheapfake” video of Nancy Pelosi went viral in February 2020 not because it was a quality forgery but because large numbers of users already held negative views of Pelosi and were happy to share them with others; a deepfake video was unnecessary.⁷⁴ Synthetic video’s largest harms to date mostly involve nonconsensual synthetic sexual imagery, a largely unsolved problem that is beginning to receive legislative attention.⁷⁵

It is too early to say with certainty if generative AI will supercharge the age of disinformation in commonly predicted ways—but it will almost certainly have important ramifications for trust and safety. These merit more methodical consideration by researchers, funders, policymakers, and technologists in order to identify the most significant risks, prioritize the most compelling solutions, and avoid the misallocation of valuable time and resources.

73 It is worth noting here that these technologies are not mutually exclusive. While forms of masking used to swap the faces of two individuals or manipulate the facial expressions of a subject are distinct from text-to-image or image-to-image models for content creation, both could be used to produce synthetic or “deepfake” images or videos.

74 Ryan, M. (2023, June 4). [AI is the New Deepfake](https://www.medium.com/@mryan1234567890/a-ai-is-the-new-deepfake-2023-06-04). *Medium*. [perma.cc/R4Y4H-3V9Q]

75 Singh, K. (2023, February 9). [There’s Not Much We Can Legally Do About Deepfake Porn — Yet](https://www.refinery29.com/en-us/2023/2/9/deepfake-porn-legal). *Refinery29*. [perma.cc/LA65-PRYD]

Political Retaliation is a Growing Threat to Disinformation Research

One of the largest threats hanging over independent counter-election-disinformation researchers in the United States is the political right's intensifying political hostility to their work. While conservative concerns about alleged social media platform bias against their viewpoints are longstanding, the past few years have seen an increase in both the volume and intensity of rhetorical, legislative, legal, and other reprisals to counter-disinformation work. While these attacks were initially aimed at platforms, they have gradually spread to government initiatives and independent researchers. In their latest manifestation, these attacks led to the injunction in *Missouri v. Biden*, which constrained the ability of the government to interact with platforms or researchers on most subjects.

This development is a serious and growing obstacle to the field. Some researchers interviewed for this report feared that orchestrated, politically motivated attacks on the field threaten to drive young professionals away, jeopardize partnerships with government agencies, spook potential funders, discredit the concept of “counter-disinformation” in the public imagination—and ultimately threaten the integrity of the 2024 elections.

Conservative Backlash Builds Through 2020 to Today

Claims of platform bias against conservatives predate the 2020 election: For example, in 2019 Facebook released the results of a “conservative bias audit” led by former Republican Senator Jon Kyl which found no concrete evidence of bias.⁷⁶ But these claims

76 Fischer, S. (2019, August 20). Exclusive: The results from Facebook's conservative bias audit. *Axios*. [perma.cc/98TE-NYEJ]

only intensified as platforms struggled to control false claims related to the election and the COVID-19 pandemic. In May, for instance, Twitter added a fact-checking label to a tweet from President Donald Trump claiming that mail-in voting would lead to widespread fraud. The response was swift and furious, leading to death threats against then-Twitter Head of Site Integrity Yoel Roth.⁷⁷ Two days later, the White House signed an executive order intended to weaken social media platforms' protection from legal liability for user-generated content.⁷⁸ A few months later in the weeks before the 2020 election, Twitter blocked a New York Post story related to Hunter Biden—a move that enraged conservatives and about which Roth would later testify in front of Congress, admitting that the company did so under the erroneous belief that the story was part of an influence operation by a foreign government.⁷⁹

Following the January 6th insurrection, Twitter, Facebook, and YouTube suspended President Trump's account, prompting a legislative response from Republican-controlled state legislatures. In May 2021, Florida Governor Ron DeSantis signed Senate Bill 7072, which allows candidates for public office to sue social media platforms if their accounts are suspended. Months later, in September, Texas Governor Greg Abbott signed House Bill 20, which prohibited platforms from “censor[ing] users or content based on viewpoint.”⁸⁰

77 Oremus, W. (2020, May 28). [Inside Twitter's Decision to Fact-Check Trump's Tweets](https://www.medium.com/@woremus/inside-twitter-s-decision-to-fact-check-trumps-tweets). *Medium*. [perma.cc/J9SU-EM6A]

78 CDT filed a lawsuit challenging this executive order: Center for Democracy & Technology. (2020, June 2). [CDT Suit Challenges President's Executive Order Targeting First Amendment Protected Speech](https://www.cdt.org/press-releases/cdt-suit-challenges-presidents-executive-order-targeting-first-amendment-protected-speech/) [Press release]. [perma.cc/8MJA-DM42]; The executive order echoed longer-standing conservative threats to dismantle the liability shield granted to internet companies by Section 230 of the Communications Decency Act. In May 2021, President Biden rescinded the order. Allyn, B. (2020, May 28). [Stung By Twitter, Trump Signs Executive Order To Weaken Social Media Companies](https://www.npr.org/2020/05/28/828111100/stung-by-twitter-trump-signs-executive-order-to-weaken-social-media-companies). NPR. [perma.cc/8ZBD-HFXU]; Lyons, K. (2021, May 15). [Biden revokes Trump executive order that targeted Section 230](https://www.theverge.com/2021/5/15/22361111/biden-revokes-trump-executive-order-that-targeted-section-230). *The Verge*. [perma.cc/8HLD-ENCZ]

79 Amiri, F. & Ortutay, B. (2023, February 8). [Ex-Twitter execs deny pressure to block Hunter Biden story](https://www.associatedpress.com/2023/02/08/ex-twitter-exec-denies-pressure-to-block-hunter-biden-story/). *Associated Press*. [perma.cc/RT5N-DLAY]

80 Brannon, V.C. (2022, September 22). [Free Speech Challenges to Florida and Texas Social Media Laws](https://www.congressionalresearchservice.org/free-speech-challenges-to-florida-and-texas-social-media-laws/). Congressional Research Service. [perma.cc/PD4W-KG6B]; Both the Florida and Texas laws were later challenged by legal suits and, as of May 2023, are under review by the Supreme Court.

Election disinformation about 2020 is fueling efforts to undermine the 2024 election—and so, counter-disinformation efforts have attracted ire and attention from election deniers.

Meanwhile, the insistence that Donald Trump defeated Joe Biden in the 2020 election became an important litmus test in Republican politics. It was arguably the key issue in several important Republican primaries in competitive states; even more moderate Republicans felt pressured to criticize the election's integrity if not to deny its outcome outright. For instance, Ohio Secretary of State

Frank LaRose—who administers elections in a state that Trump won by a significant margin—appeared on a panel called “They Stole It From Us Legally” at the 2023 Conservative Political Action Conference (commonly called by its acronym, CPAC).⁸¹

It appears that conspiracy theories about election integrity in 2020 will, ironically, have a substantial negative impact on the integrity of upcoming elections. Since 2013, the Electronic Registration Information Center (ERIC) has been a consortium

enabling states to maintain accurate voter rolls by sharing information about individuals who die or move across state lines. But narratives implying that ERIC was a “left-wing plot to register Democratic voters and steal elections” have led to the withdrawal of several large Republican-led states, including Ohio, Virginia, Florida, and Texas.⁸²

Election disinformation about 2020 is fueling efforts to undermine the 2024 election—and so, counter-disinformation efforts have attracted ire and attention from election deniers. After the 2020 election conservative activists ratcheted up rhetorical and legal attacks against independent researchers: In September 2021, the same month the Texas bill was signed into law, conservative activist

81 BeMiller, H. (2023, March 6). Secretary of State Frank LaRose touts Ohio elections alongside election deniers at CPAC. *The Columbus Dispatch*. [perma.cc/EC3F-X3GF]

82 Parks, M. (2023, June 6). How the far right tore apart one of the best tools to fight voter fraud. NPR. [perma.cc/AR5M-9YUD]; Paviour, B. & Parks, M. (2023, May 11). Virginia becomes the latest GOP-governed state to quit a voter data partnership. NPR. [perma.cc/H3D7-LP99]; Contreras, N. (2023, July 20). Texas resigns from ERIC, a national program that keeps voter rolls updated. *VoteBeat*. [perma.cc/5RMH-LWGV]

group Project Veritas filed a defamation lawsuit against Stanford University and the University of Washington, which host two members of the Election Integrity Partnership who had criticized Project Veritas's work in a blog post.⁸³ As a public institution, the University of Washington must comply with the Freedom of Information Act (FOIA); after the EIP submitted a statement to the Select Committee to Investigate the January 6th Attack on the United States Capitol, the University received a large volume of FOIA requests from conservative figures and media outlets.⁸⁴ Among the records requested were all communications with social media platforms and a variety of federal agencies and officials.⁸⁵

The Short Life of the Disinformation Governance Board Marks a Turning Point

Individuals affected by these trends said in interviews for this report that they view April 2022 as an inflection point for these intimidation tactics. That month, the Department of Homeland Security (DHS) announced the creation of a new body, the Disinformation Governance Board, to coordinate responses to misinformation with homeland security implications. Nina Jankowicz, a prominent researcher, was tapped to run the organization. The response from conservative media was again swift: activists decried the board as a “ministry of truth,” it received condemnation from

- 83 Project Veritas later lost the suit and was required to pay Stanford's legal fees; ultimately, for the researchers involved, it was more a drain of time than financial resources. Masnick, M. (2022, August 8). [Project Veritas Not Only Loses Its Vexatious SLAPP Suit Against Stanford, It Has To Pay The University's Legal Fees.](https://perma.cc/4FGW-DWMV) *Techdirt*. [perma.cc/4FGW-DWMV]
- 84 University of Washington professor Kate Starbird discussed her experience going from an observer of conspiracy theories to the subject of them here: Magby, J. (Host.) (2023, February 24). [Post-Election Audits, Disinfo — Talking Tech w/ Kate Starbird, Will Adler, Aliya Bhatia](https://perma.cc/333M-L2LD) [Audio podcast episode]. In *CDT's Tech Talk*. Center for Democracy & Technology. [perma.cc/333M-L2LD]
- 85 Dudley, B. (2022, October 19). [Harassment, public-records requests bombard UW truth seeker after Jan. 6 hearings came.](https://perma.cc/CXC5-6XJP) *The Seattle Times*. [perma.cc/CXC5-6XJP]

conservative lawmakers, and Jankowicz herself was attacked across conservative media; her image even appeared on the cover of the *New York Post*.⁸⁶ After three weeks of intense criticism, the Biden administration dissolved the Board and Jankowicz resigned.

In an interview for this report, Jankowicz said the Board was meant to be a coordinating body at DHS which kept different teams from working at cross purposes. Given the risk that the Board could be misconstrued as something more sinister, Jankowicz said she encouraged DHS to be as transparent as possible by making a multi-pronged announcement, including a press release with a factsheet and a pre-briefing for media and Congressional staff. But bureaucratic obstacles prevented this approach, and in a relative void of information, far-right actors seized on the Board and began a campaign against it.

The result was a barrage of negative publicity driven primarily by conservative media and Congressional Republicans that Jankowicz said drove harassment toward her and her family.⁸⁷ She estimates that at one point, her face was on Fox News once an hour.⁸⁸ During this period, Jankowicz became concerned for her safety. She lacked confidence that DHS would be able to identify threats against her, so she hired a private online security consultant to monitor the “dark web” for discussions about her.

- 86 Lorenz, T. (2022, May 18). How the Biden administration let right-wing attacks derail its disinformation efforts. *Washington Post*. [perma.cc/GVF4-38F4]; Myers, S.L. & Sullivan, E. (2022, July 6). Disinformation Has Become Another Untouchable Problem in Washington. *New York Times*. [perma.cc/RP4J-F8KY]; The Editorial Board. (2022, April 28). Which useful idiot thought such a clearly partisan hack should be Biden’s ‘disinformation czar’? *New York Post*. [perma.cc/8J9N-4KUA]
- 87 More tempered, and less personalized, were concerns and criticisms from civil rights groups. See: Protect Democracy, Electronic Frontier Foundation, & Knight First Amendment Institute at Columbia University. (2022, May 3). Letter to Secretary Alejandro Mayorkas re: Significant Concerns Regarding the “Disinformation Governance Board.” [perma.cc/QVU4-W5XZ]
- 88 Kotsonis, S. & Chakrabarti, M. (2023, May 15). What happened to Nina Jankowicz when Fox News came for her [Audio broadcast]. In *On Point*. WBUR. [perma.cc/A8J3-ZYX2]



Figure 10. Representative Lauren Boebert (R-CO) making a statement about Jankowicz on the House floor.

Representative Boebert is standing next to a poster with lyrics from the Cabaret song “My Simple Christmas Wish,” which Jankowicz performed as part of a community theater performance.



Source: *Forbes Breaking News*. (2022, May 10). [‘Okay, Let’s Give Them Something To Talk About...’: Lauren Boebert Rips Biden ‘Ministry Of Truth’ \[Video\]](#). YouTube. [perma.cc/EQ7M-TRX7]

The Friday after her position was announced, the consultant told her that the pitch of online rhetoric about her was so alarming that she should leave her home. Because Jankowicz was still formally employed by DHS and it is a crime to threaten a federal official, the Federal Protective Service was eventually involved. It sent dozens of subpoenas to social media platforms to unmask individuals who made threats against her online.

Jankowicz said that things did not calm down after her resignation from DHS. Fox continued covering her when she became a private citizen—by one count she was mentioned on the channel more than 300 times in eight months, a period of time much longer than her tenure at DHS.⁸⁹ “Every mention brought in a wave of harassment,” she said. Jankowicz has obtained protective orders against a stalker who, claiming to be a journalist, doxxed her; Jankowicz was concerned he would film and livestream himself outside of her home. She has been recognized in public, and for a time took to wearing a medical facemask and a hat when leaving the house. The day of her interview for this report, more than a year after the Board was announced, Jankowicz received a Google Alert that someone had used synthetic “deepfake” video technology to produce and disseminate nonconsensual sexual images with her face.⁹⁰

89 Falconer, R. (2023, May 10). [Former Biden admin disinformation chief sues Fox News for defamation](#). *Axios*. [perma.cc/NG7F-49KF]

90 Jankowicz, N. (2023, June 25). [I Shouldn’t Have to Accept Being in Deepfake Porn](#). *The Atlantic*. [perma.cc/6JJT-M8A4]

One researcher described the concerted attacks that came in the aftermath of the Disinformation Governance Board's dissolution as an "effort to create a Pavlovian response to the word 'disinformation,'" allowing political opponents to use conspiracy theories to pressure researchers' potential partners.

Many individuals interviewed for this report saw the Disinformation Governance Board debacle and continued vilification and harassment of Jankowicz as an early proof-of-concept for today's attacks on the field. One researcher described the concerted attacks that came in the aftermath of the Disinformation

Governance Board's dissolution as an "effort to create a Pavlovian response to the word 'disinformation,'" allowing political opponents to use conspiracy theories to pressure researchers' potential partners. A former platform staffer said that politicization of counter-disinformation following the announcement of the Board problematized otherwise helpful DHS efforts to work with industry to stay atop disinformation threats.

Women and people from marginalized groups face especially acute risks. An individual affiliated with the DDL said in an interview that DDL members often talk about harassment: while they are somewhat insulated from it by the nature of DDL's closed, non-public membership, they are aware of others in the field, especially Black women and journalists, who face this problem.

Disinformation researchers worry that experiences like this will turn young professionals away from their field, and Jankowicz fears the effect will be especially pronounced for young women. "If a man, even a young man, were in the post [of Executive Director of the Board], I don't think the ridicule would have been as juicy," she said, noting that she was sexualized in the media and even by Members of Congress. She said that young women interested in her work have asked her about this before—they see harassment against other prominent women in public life and want to avoid similar treatment, even at the cost of curtailing their self-expression and professional ambition.

The Twitter Files and the Select Committee on Weaponization of the Federal Government Bring Challenges to a Boiling Point

The misrepresentation of the “Twitter Files” and the subsequent Congressional hearings have heavily politicized relationships between researchers, platforms, and the government.

The sum of these events is that by the time the Republican Party won control of the House of Representatives in 2022, the conservative campaign against counter-disinformation had gained serious momentum. It was boosted further by Elon Musk’s purchase of Twitter that October. In the waning days of 2022, Musk released a batch of Twitter correspondence between platform staff, law enforcement, and outside researchers from the 2020 election (the

“Twitter Files”) to a group of writers who alleged that federal agencies and the Biden administration pressured Twitter into removing conservative speech during the 2020 election, sometimes with the assistance of university and nonprofit researchers.⁹¹

The misrepresentation of the “Twitter Files” and the subsequent Congressional hearings have heavily politicized relationships between researchers, platforms, and the government. However, the allegations made by their promoters include major inaccuracies. Some of the targeted institutions have responded publicly. Stanford, for instance, was accused of labeling 22 million tweets as “misinformation”; the University clarified that it analyzed 22 million tweets, found fewer than three thousand to be in violation of Twitter’s stated policies, and merely informed the company of those tweets without making recommendations toward what actions Twitter should take independently.⁹² The University of Washington likewise put out a statement, correcting, among other “false impressions,” claims that it served as a way to “route”

91 For more on the Twitter Files, see: Coldewey, D. (2023, January 13). [Deconstructing ‘The Twitter Files.’ *TechCrunch*. \[perma.cc/K7DU-ZC5M\]](https://perma.cc/K7DU-ZC5M)

92 Stanford Internet Observatory. (2023, March 17). [Background on the SIO’s Projects on Social Media. \[perma.cc/Z27L-G6TV\]](https://perma.cc/Z27L-G6TV)

content moderation requests from government and other external actors to platforms, and that it was designed as a “cut-out” from the Cybersecurity and Infrastructure Security Agency (which was led by a Trump administration appointee during the 2020 election).⁹³

The communications between the government and platforms were also broadly misrepresented—no evidence in the files suggests the government (which, again, was at the time led by the Trump administration) used the force of law or other measures to compel or coerce platforms to remove or filter conservative content, and platforms did not act on all of the concerning content flagged for them by government agencies. In fact, the released internal emails from Twitter show measured and contentious debate between teams and senior staff about content moderation decisions.⁹⁴ Many claims about the contents of the files have been demonstrably false—to the extent that Twitter’s own lawyers refuted many of them in court when President Trump sued Twitter over the suspension of his account.⁹⁵

Regardless, the files were instantly seized upon by Representative Jim Jordan, a Republican from Ohio, who vowed to investigate them as chairman of both the House Judiciary Committee and the new Select Committee on the Weaponization of the Federal Government. Jordan argued that the FBI and other federal bodies had pressured or otherwise colluded with social media platforms and third-party researchers as part of a “Censorship Industrial Complex” against conservatives during the 2020 election.⁹⁶ He held

-
- 93 Starbird, K., Calo, R., Coward, C., Spiro, E.S., & West, J.D. (2023, March 16). Addressing false claims and misperceptions of the UW Center for an Informed Public’s research. *University of Washington Center for an Informed Public*. [perma.cc/RD2B-A7MA]
- 94 Duffy, C. (2022, December 14). The real revelation from the ‘Twitter Files’: Content moderation is messy. *CNN*. [perma.cc/H2Y7-JSDF]
- 95 Twitter, Inc., et al. (2023, June 1). DEFENDANTS’ OPPOSITION TO PLAINTIFFS’ MOTION FOR INDICATIVE RULING. U.S. District Court Northern District of California San Francisco Division. [perma.cc/EP3V-ZCUX]
- 96 U.S. House of Representatives Judiciary Committee. (2023, January 10). Jim Jordan on Why the Select Subcommittee on the Weaponization of the Federal Government is Necessary [Press release]. [perma.cc/2MHE-VLAV]; For Congressional testimony related to these claims, see: Shellenberger, M. (2023, March 9). Testimony on “The Censorship Industrial Complex: U.S. Government Support For Domestic Censorship And Disinformation Campaigns, 2016 - 2022” before the U.S. House of Representatives, House Select Committee on the Weaponization of the Federal Government. [perma.cc/QX7T-Y9HV]

a Congressional hearing with former senior Twitter staff (like Yoel Roth, as mentioned previously), Matt Taibbi, and other witnesses; he also subpoenaed several research initiatives, including members of the EIP, the German Marshall Fund, and Clemson University.⁹⁷

On June 1st, 2023, Jordan sent a letter to Stanford warning that its compliance with the subpoena was “insufficient” and that if Stanford did not produce more documents, the Judiciary Committee “may be forced to consider the use of one or more enforcement mechanisms.”⁹⁸ The next month, Twitter filed a lawsuit against the Center for Countering Digital Hate (CCDH) over a study of hate speech on the platform, and the House Judiciary Committee announced plans to investigate CCDH as well.⁹⁹

The more efforts to recast counter-election-disinformation as censorship succeed, the more difficult it will become for governments and others to work with researchers in this field.

In interviews, counter-disinformation professionals raised concerns that these developments will have a “chilling effect” across the field. The more efforts to recast counter-election-disinformation as censorship succeed, the more difficult it will become for governments and others to work with researchers in this field: government or foundation employees, for example, could become unwilling to so much as share links to researchers’ work for fear those messages will become public and be used against them. Many professionals in this space now operate more carefully because they are aware that at any time there could be a lawsuit (or, at public institutions or those that work with government partners, a FOIA request), which might lead to the public release of their email communications. Even if these messages reveal nothing scandalous, professionals are

- 97 Myers, S.L. & Frenkel, S. (2023, June 19). [G.O.P. Targets Researchers Who Study Disinformation Ahead of 2024 Election](#). *New York Times*. [perma.cc/ZYZ7-UTFJ]; Bernstein, A. (2023, March 22). [Republican Rep. Jim Jordan Issues Sweeping Information Requests to Universities Researching Disinformation](#). *ProPublica*. [perma.cc/3T2Y-EHCC]
- 98 U.S. House of Representatives Judiciary Committee. (2023, June 1). [Chairman Jordan Presses Stanford on Subpoena Compliance for Censorship Investigation](#) [Press release]. [perma.cc/ML92-A9ND]
- 99 Cristiano, L. (2023, August 15). [Under fire from Musk and the GOP, nonprofit chief vows to forge ahead](#). *Washington Post*. [perma.cc/KE2H-ZNCX]

increasingly aware that they can be used out of context to create the appearance of scandal. Some said they now do more work by phone, and less by email. Even if they are merely engaged in information sharing, researchers try harder than ever to avoid giving officials reasons to keep that at arm's length.

The manufactured scandals around disinformation research have already damaged researchers' relationships with platforms. Social media companies have also been subpoenaed by the House Judiciary Committee, making platforms more cautious about external communications. One researcher said that the "level of candidness and trust" with corporate counterparts is gone because any interaction could become a potential scandal. "Since 2016, the project has been to build relationships between CSOs, companies, and law enforcement," they said; "It had gotten to a place where it was working, and now it's ten steps back." In this environment, communication between researchers and platforms could become so sanitized as to lose value.

The Outcome of *Missouri v. Biden* is Uncertain but Potentially Debilitating

In May 2022, the Attorneys General of Missouri and Louisiana filed a lawsuit, *Missouri v. Biden*, against the Biden administration for "allegedly colluding" with social media companies in order to "censor and suppress free speech."¹⁰⁰ The district court judge overseeing the case issued an injunction on July 4, 2023, forbidding significant parts of the federal government from communicating with platforms and independent researchers to monitor and respond to social media

100 Missouri Attorney General Eric Schmitt. (2022, May 5). [Missouri, Louisiana AGs File Suit Against President Biden, Top Admin Officials for Allegedly Colluding with Social Media Giants to Censor and Suppress Free Speech](#) [Press release]. A similar class action suit, *Hines et al v. Stamos et al.*, has been filed against both individuals and institutions involved in the EIP and is also discussed in the Stanford amicus brief: [Brief of amici curiae Stanford University, Alex Stamos, and Renée DiResta in Support of Appellants, *Missouri v. Biden* \(2023\)](#). [perma.cc/A4T5-D7ZQ]

content containing protected speech.¹⁰¹ The 5th U.S. Circuit Court of Appeals subsequently narrowed the injunction and as of publication, the order has been stayed.¹⁰² But if the order goes back into effect in some form, it could dramatically alter the landscape for counter-disinformation in the 2024 election.¹⁰³

Several of the groups studied for this report were mentioned by name in the lower court’s injunction, and Stanford filed an amicus curiae brief on appeal in the case explaining that the injunction falsely attributed statements to individual Stanford researchers.¹⁰⁴ The lower court injunction specifically forbade the government from “collaborating, coordinating, partnering, switchboarding, and/or jointly working with the Election Integrity Partnership, the Virality Project, the Stanford Internet Observatory, or any like project or group for the purpose of urging, encouraging, pressuring, or inducing in any manner removal, deletion, suppression, or reduction of content posted with social-media companies containing protected free speech.”

The Fifth Circuit reversed the portion of the injunction that applied to independent researchers, but it remains to be seen what happens as the case proceeds. Moreover, the fact that the lower court enjoined many communications with independent researchers about content moderation may well have a chilling effect on such communications.

101 *State of Missouri, et al. v. Joseph R. Biden, Jr., et al.*, No. 3:22-CV-01213. (U.S. District Court, Western District of Louisiana, Monroe Division. (Order granting preliminary injunction.) [perma.cc/PUF2-2DG4]

102 *State of Missouri, et al. v. Joseph R. Biden, Jr., et al.*, (5th Cir. Sept. 8, 2023). <https://www.ca5.uscourts.gov/opinions/pub/23/23-30445-CV0.pdf>

103 Myers, S.L. & McCabe, D. (2023, July 4). Federal Judge Limits Biden Officials’ Contacts With Social Media Sites. *New York Times*. [perma.cc/S25B-6QWW]

104 Brief of amici curiae Stanford University, Alex Stamos, and Renée DiResta in Support of Appellants, *Missouri v. Biden* (2023). [perma.cc/A4T5-D7ZQ]

A sweeping ban against counter-disinformation coordination between stakeholders is harmful; more constructive would be transparency around communications between government officials, platforms, and outside researchers.

While most disinformation researchers are troubled by these developments, some of the individuals interviewed for this report expressed sympathy for concerns about government policing of speech and the idea that it should be more transparent in its interactions with social media platforms. The fear of government pressure leading to social media censorship, while not validated by

the Twitter Files, is not without basis internationally: the Indian government, for instance, passed a law in 2021 requiring platforms to have a representative in-country who could be imprisoned for failing to comply with content takedown requests. Twitter has challenged the law in court; the suit is ongoing.¹⁰⁵ The Indian government seems undeterred: In April 2023, it amended legislation to require social media platforms to adhere to the findings of a government-run fact-checking unit.¹⁰⁶

A sweeping ban against counter-disinformation coordination between stakeholders is harmful; more constructive would be transparency around communications between government officials, platforms, and outside researchers, like those included in the Twitter Files. For example, platforms could report government communications of their own accord, similar to how they report government requests for personal data around the world.¹⁰⁷

105 Singh, K.D. & Conger, K. (2022, July 8). [Twitter, Challenging Orders to Remove Content, Sues India's Government](https://www.nytimes.com/2022/07/08/technology/twitter-challenges-india.html). *New York Times*. [perma.cc/AJ2Q-3F3V]

106 Singh, M. (2023, April 6). [India to require Facebook and Twitter rely on gov't fact checking](https://www.techcrunch.com/2023/04/06/india-to-require-facebook-and-twitter-rely-on-govt-fact-checking/). *TechCrunch*. [perma.cc/T43F-5XUE]

107 For example, see: Meta. (n.d.). [Government Requests for User Data](https://www.meta.com/privacy/government-requests). [perma.cc/5FZ2-N5QW]

Funding Challenges Cloud the Horizon

Going into 2024, a confluence of trends—some political, some economic—has complicated the funding prospects for professionals in this field. While the effects have not been dire yet, many disinformation researchers are bracing for a contraction. They reflected on this in interviews for this report. Platform staff, on the other hand, offered their perspective on how fundraising incentives have complicated past efforts.

One basic fundraising obstacle is the perception among funders, based on the 2022 elections, that dangers to U.S. democracy and election disinformation, in particular, may be on the down slope.

Jesse Littlewood at Common Cause noted that donors can be reactive, and it can be difficult for civil society to prove it prevented a disaster from happening. He called 2022 the kind of victory that creates a funding challenge, saying that “you do not raise money off a victory” and that funders should make long-term investments in infrastructure to more durably reduce the threat.

Unfortunately, the political dynamics outlined earlier in this report also threaten funding for counter-election-disinformation initiatives. Littlewood said that this backlash is a signal the work is effective, not a reason to pull back—but noted a clear chilling effect on civil society organizations for whom dealing with political backlash would be a resource drain and a distraction from serving constituent communities. Others interviewed for this report fear that, wishing to avoid backlash from conservative media and politicians, government agencies and even foundations could decrease funding for counter-disinformation initiatives. This would damage the sustainability of the field, leading to scarcer funding overall and more reliance on national security-related donors (who are somewhat insulated from

One basic fundraising obstacle is the perception among funders, based on the 2022 elections, that dangers to U.S. democracy and election disinformation, in particular, may be on the down slope.

political criticism) or the most liberal or progressive foundations.¹⁰⁸ A researcher interviewed for this report said some grants have already fallen through, with funders backing away apologetically or without explanation.

Platforms are also a funding source for independent disinformation research, though a controversial one given that they are often the subject of that work. But given layoffs and other economic trends in the tech sector, this type of funding seems likely to shrink as well. Some researchers fear that grant funding could become more competitive as external initiatives currently financially supported by platforms are forced to find other sources of funding. Even if platforms were to increase their support, it would present the field with a conundrum: as explained earlier in this report, the provision of free content moderation for platforms by civil society can be seen as extractive, so compensation would appear to be a solution. But compensation from the subjects of accountability research can just as quickly turn into a form of corporate capture of outside watchdogs.

Considering the political and economic dynamics facing disinformation research, foundations and other non-government sources of support outside the corporate sector appear to be the best funding prospects for the field. Platform staff suggested these funders should create more stable, longer-term grants to accountability organizations and researchers in order to relax what they saw as negative funding incentives in the field: though not a reflection on any specific group or the groups discussed in this report, in interviews platform staff expressed belief that the need to raise grant money sometimes causes nonprofit and university researchers to prioritize press attention over long-term impact in order to attract funding. While advocacy groups interviewed for this report contest that claim, the recommendation for longer-term, less project-based funding has consensus support, echoing other analysis of the field.¹⁰⁹

108 In the words of the researcher quoted at the top of this paragraph, “the national security world runs on its own clock,” and may be insulated from partisan criticism in a way other government entities are not.

109 Goodwin, C. & Jackson, D. (2022, February 9). [Global Perspectives on Influence Operations Investigations: Shared Challenges, Unequal Resources](https://perma.cc/H372-QBAF). *Carnegie Endowment for International Peace*. [perma.cc/H372-QBAF]

04

Adapting for 2024 and Beyond

Researchers share common fears that the kind of election integrity efforts that were possible in 2020 or even 2022 will not be possible going forward as organizations decide “they want nothing to do with this.” But they insisted that the threat is only getting larger and that the field needs to grow, not shrink. Many pointed to layoffs on platform trust and safety teams and fears about generative AI as signs of worse to come. For others, these concerns are secondary to the normalization of disinformation. Nina Jankowicz said that for her, the most visceral moment of 2023 so far had been the attack on Paul Pelosi, husband of the former Speaker of the House of Representatives, in his home that January, and subsequent lies and rumors about the attacker and his supposed relationship with Pelosi. “An eighty-year-old man was attacked with a hammer and people politicized it,” she said.

Previous approaches to independent election disinformation research arguably left much to be desired in better conditions than society faces today. For the 2024 U.S. elections and whatever lies beyond, researchers need to revise the theory of change behind their work while navigating treacherous political waters and with fewer partners in the technology sector.

What do Disinformation Researchers Need in a Hostile Environment?

Prominent researchers may face strategic lawsuits against public participation (or SLAPPs), investigations, harassment, and threats of violence. Addressing these takes time, expertise, and perhaps most significantly, resources. In interviews, professionals in this field urged institutions to think about “what will happen when this comes for them.” How will they support their staff? What types of security, legal, communications, and even moral support might their staff need? Unfortunately, in the past many targeted individuals have been left to fend for themselves.

Researchers Need Security, Online and Off

Nina Jankowicz says that security costs like those she incurred don't decrease with time. On the contrary, for her, they have compounded. Faced with ongoing harassment, Jankowicz paid out of pocket for a home security camera and other security measures. She described the assistance she received from DHS as wanting—the Department did not monitor key internet forums like 8kun for threats against her and when she said she felt unsafe at home, she was told she could work from the office more often. That, of course, offered no protection outside working hours, and, in any case, Jankowicz was pregnant at the time, and COVID-19 infection rates were high.

Many researchers take steps to improve their digital and physical security—for example by hiring firms to find and remove their personal information from the internet. These practices need to be adopted in advance of targeted harassment—once harassers find an address, it's too late. Institutions should think well in advance about threats to their researchers and how to mitigate them. In at least one example, an institution subpoenaed by the House Judiciary Committee received grant support for online threat monitoring.

Legal Harassment Brings High Costs

Legal fees are also a challenge—both financially and as a drain on researchers' time and energy. Jankowicz personally covered legal costs arising from her harassment, including a restraining order against a man who filmed himself outside her home. Legal costs even challenge researchers in larger institutions: an individual who asked to remain anonymous said that after they were subpoenaed by the House Judiciary Committee, they were lucky to receive pro bono legal assistance through a personal connection of their leadership. "Nobody knew what to do from a legal perspective," they said. "Our entire legal strategy was, 'talk to every lawyer you have a connection with.'"

This interviewee said their team was fortunate to have informal networks to draw on; independent researchers or smaller teams might not, and the expense of paying for lawyers to help with subpoena compliance and testimony out of pocket could easily be "catastrophic." In the future, formal networks of attorneys who can provide this kind of pro bono support could support less well-connected researchers.

Crisis Communications is a Major Gap

Public relations strategies were also a major theme in conversations about challenges facing disinformation researchers today. Many who have faced public rhetorical attacks offered advice based on their experience. They said that risk-averse institutions like government departments, universities, and large nonprofits often follow a traditional playbook of laying low and not responding.

This approach may help weather a bad news cycle, but against coordinated, sustained activist attacks it can be counterproductive. Non-response allows attackers to control the narrative and deprives researchers of opportunities to refute inaccurate portrayals of their work. Jankowicz said after she and the DHS came under attack, the Department was "totally stuck, like a deer in headlights." She called the response from DHS and the Biden administration "dithering"

and “muted,” recalling that it took weeks for DHS to release a pre-drafted fact sheet that rebutted many of the accusations against DHS. Abuse and falsehood filled the vacuum of the Department’s non-response.

Other researchers report similar experiences, saying their institutions “misunderstood what they were facing from the beginning” and that they had to fight for a more proactive public response—like creating fact sheets that credible journalists can reference when reporting on controversies. “You can’t turtle silently,” said an anonymous researcher. “You will lose control of the narrative and be forced to respond at various points... if you wait, the questions will be dictated by someone else.” This is a strategy best adopted early: once a narrative reaches mass audiences, a delayed response may look defensive and backfire by drawing older claims back into the news.

Stanford researcher Renee DiResta’s public account of her interactions with Twitter Files writer Michael Shellenberger is an instructive example. After DiResta felt Shellenberger mischaracterized their correspondence in his Congressional testimony and later media appearances, she released their texts and emails so the public could judge for themselves.¹¹⁰ In the words of the anonymous interviewee, “Let’s not play by normal rules if a journalist is going to lie. Point out inaccuracies, embarrass him a little bit, make people question the reporting at large.” In a similar tactic, following Shellenberger’s Congressional testimony, the German Marshall Fund submitted a correction to the Congressional record describing Shellenberger’s statements as “incorrect.”¹¹¹

110 DiResta, Renée. (2023, March 31). [Fiction vs Reality: My Texts with Michael Shellenberger](#). *Renee’s Substack*. [perma.cc/7UMW-69VC]

111 Salvo, D. & Wilson, R.D. (2023, March 10). [Letter to Representatives of the U.S. House Judiciary Committee](#). *The Klonickles* [perma.cc/T65F-Z3FJ]

Psychological Support is Important

Increased mental healthcare and community moral support are two final forms of assistance researchers agree they need. Disinformation researchers are human, and these experiences have a psychological toll. Researchers say that access to therapy is critical. They also said the community as a whole should be more vocal in defense of peers who are facing these challenges. Nina Jankowicz said that when the harassment against her began, “most of the counter-disinformation community said nothing.” She attributes this to a mix of reasons, from disapproval of DHS to fear that they would be drawn into the harasser’s line of fire. The experience left her feeling isolated and, due to DHS policy, unable to speak up for herself. This is the position that many researchers today are seeking to avoid through more open and proactive communication. But, some still feel they lack sufficient channels to discuss these issues as a professional community or to form a united front.

Having Soured on Platform Partnerships, Many Initiatives Are Exploring Other Approaches

The decline of trust and safety has left many independent counter-election-disinformation efforts to reconsider their approach to improving platform content moderation. As they prepare for the 2024 election, many are shifting gears.

A critical observer might say that these responses indicate an end to the counter-disinformation field’s ambition for a systemic solution—that they reflect Jankowicz’s fears that disinformation has been normalized, and society is resigned to it. But interviewees for this project disagreed. They hoped that, through continued research, the field could inform strategies and policy changes that will eventually help shape a healthier information environment.

Some Organizations are Doubling Down on Pre-bunking and Counterspeech

One researcher said that the field is largely giving up on “piece-by-piece” content moderation and exploring alternatives. A researcher at Common Cause acknowledged that it might be impossible to do this work with sufficient speed or at a large enough scale: no matter how many posts they convince platforms to take action against,

the next day, or minute, or hour, there will be more. In this context, counter-disinformation becomes a game of “whack-a-mole” in which it becomes difficult to measure impact. As for catching this content before it goes viral, former Facebook engineer Glenn Ellingson said in an interview that platform response time to inbound alerts is typically too slow to respond early in the life cycle of a post, when it can prevent the most spread. Disinformation monitoring is almost certainly more valuable for its ability to provide situational awareness, shape counter-messages, and inform policy recommendations for platforms and

government. But, as a staffer at the Disinfo Defense League said in an interview, even policy advocacy aimed at platforms can be “ephemeral”: They cited Twitter as an example, pointing to dramatic layoffs and Elon Musk’s repeated rollbacks of years of integrity policy improvements.

In an interview, a researcher at Common Cause said the organization is pivoting their interventions from focusing on reporting policy violations to platforms, citing disappointing results of those relationships in 2020 and the companies’ decreasing responsiveness since. While Jesse Littlewood said that Common Cause will continue reporting content to platforms when it is clearly in violation, Common Cause may invest its limited time and resources more heavily in counterspeech and the provision of accurate information through grassroots organizations. For 2024, it plans to produce toolkits to help “hammer home” messages about,

Disinformation monitoring is almost certainly more valuable for its ability to provide situational awareness, shape counter-messages, and inform policy recommendations for platforms and government.

for example, the security of voting by mail. Common Cause tried this approach in 2020 with some success, when it encouraged its network of volunteers to help spread messages supporting voting and “pre-bunking” election disinformation.¹¹² As evidence of effectiveness, one Common Cause staffer cited polling showing the public understood that delays in counting mail-in and early votes meant “election night [would not be] results night,” a message jointly and repeatedly emphasized by election protection groups.

Others Focus on Informing Election Officials

A researcher affiliated with the Institute for Strategic Dialogue (ISD) said that there has been a conscious movement in the field away from partnerships with platforms—motivated in part by questions about how genuinely social media companies care about election issues. They said that platforms are “not really listening in the way we need them to,” and that they do not act quickly or consistently enough to make these partnerships effective. While quick to stress that there are “good people” in the companies who do care about election integrity, they also noted that “in the grand scheme of things, these are huge corporations” whose decisions are beholden to shareholders and profit margins.

Instead of working through platforms, the ISD researcher said that the logical next place for counter-disinformation to focus is on people who are impacted directly by false claims about elections. ISD has focused its efforts on liaising with election officials, especially those in states that have become “hot spots” for election disinformation and extremist activity. ISD uses those relationships to help state governments have an informed response—alerting those governments to threats against their staff and teaching them to distinguish serious threats from elevated rhetoric.

112 Bond, S. (2022, October 28). [False information is everywhere. 'Pre-bunking' tries to head it off early.](https://www.npr.org/2022/10/28/1111111111) *NPR*. [perma.cc/Z5PV-L5FC]

Many Initiatives Seek Policy Advocacy and Other Levers to Change Negative Incentives in the Media Environment

Some researchers never saw it as their goal to work with platforms to improve content moderation. A staffer for the DDL said that they saw relationships with platforms as “fraught” from the outset. “I don’t want to work with platforms on how to regulate platforms,” they said, adding that self-regulation has not been conducive to change. Other researchers see relationships with platforms as a “can of worms”: too close a relationship with platforms can raise uncomfortable questions about the independence of research, the appropriateness of accepting exclusive access to data or other resources, and the risk of being singled out for political attacks.

DDL is shifting from counter-messaging as a rapid response to longer-form research informing advocacy efforts and policy proposals. Informed by analysis of DDL’s strengths, weaknesses, and most valuable contributions to the field, DDL staff decided that as a small team, they are not best equipped to monitor the broader internet for racialized disinformation—though they have trained other organizations to do this work at greater scale, with more capacity. DDL now focuses on producing detailed quarterly reports analyzing what types of narrative interventions are effective in specific contexts (for example, attacks on Asian Americans during the COVID-19 pandemic). In interviews, DDL staff said they are still waiting to see what kind of advocacy efforts their research enables.

A researcher at Common Cause similarly said their best hope for a long-term solution lay not in content moderation practices, but ways of forcing better platform design choices and altering the incentives social media creates for producing and spreading disinformation. “We need a strategy that isn’t just reacting to what’s gone viral. We need to make election denial more toxic and less profitable,” they said, citing recent judgments against Trump campaign lawyers for making false claims about the election in court as an example of accountability.

Other examples from the legal system include the lawsuit between Fox News and Dominion Voting Systems and the ongoing suit between Fox and Smartmatic, another voting machine provider. Fox agreed to pay \$787.5 million in a settlement with Dominion—a tremendous sum that may act as a deterrent to future dishonesty.¹¹³ Others worry that because the case ended in a settlement, Fox avoided the worst public relations ramifications of a protracted trial and that after the ouster of former Fox News host Tucker Carlson for comments he made in text messages, the main lesson from the lawsuit will be to communicate more cautiously. But Fox is not out of legal jeopardy yet: In May 2023, Nina Jankowicz also filed suit against Fox for defamation.¹¹⁴

113 Kelley, L. (2023, April 19). [Fox News Settled Its Suit, but Similar 2020 Election Cases Are Pending](https://perma.cc/2H8E-WTJB). *New York Times*. [perma.cc/2H8E-WTJB]

114 Rutenberg, J. & Myers, S.L. (2023, May 10). [New Defamation Suit Against Fox Signals Continued Legal Threat](https://perma.cc/3C2Z-C6TM). *New York Times*. [perma.cc/3C2Z-C6TM]

05

Recommendations

Absent major reversals, the 2024 election will likely be the most favorable environment for disinformation in the United States since 2016. The electoral stakes have only increased: as Bridget Barrett and Daniel Kreiss argue in a piece for Tech Policy Press, Americans are “running out of time and chances to continue our experiment with democracy.”¹¹⁵

Against this backdrop, efforts to protect U.S. elections from disinformation are in crisis. Platforms have proven to be unreliable partners: despite vast resources and public pressure, they generally have failed to create dependable points of contact, meaningfully increase transparency, consistently act on leads to clear terms of service violations, and take proactive steps to limit the spread of harmful content. As Barrett and Kreiss lament, Meta, YouTube, and Twitter “appear to be taking a laissez-faire approach to 2024, as they did in 2016” and have done away with policies against disinformation about the 2020 and 2022 elections. This is a grave

115 Barrett, B. & Kreiss, D. (2023, June 29). [Platforms are Abandoning U.S. Democracy](https://perma.cc/2T5T-TW68). *Tech Policy Press*. [perma.cc/2T5T-TW68]

concern—false rhetoric from Trump and others about those elections primes voters for future conspiracy theories about the 2024 election and risks increased harassment of election workers.¹¹⁶ Such harassment could easily spill over into violence.

Simultaneously, the terrain for election integrity has become considerably more challenging over the roughly one year it took to produce this report. Tens of thousands of tech workers lost their jobs, including many who work on trust and safety and content moderation. An eccentric billionaire purchased one of the world's most important communications channels and undermined many of its prior efforts to combat disinformation. Researchers are weathering sustained political attacks, weary from legal and PR battles, and bereft of levers for corporate accountability. Government agencies may be barred from engaging with both researchers and platforms, and legislators have failed to pass major overhauls of internet regulation. Meanwhile, generative AI potentially threatens to exacerbate the scope and volume of disinformation.

This moment calls for a reevaluation of strategies for protecting U.S. elections.

This moment calls for a reevaluation of strategies for protecting U.S. elections. The types of partnerships between platforms, government, and researchers set up for the 2020 U.S. election may no longer be feasible. These partnerships have provided valuable insight into how false and misleading narratives about elections develop and spread. But their practical impact—at least in terms of improving content moderation and reforming platform policy—has been dampened by both the magnitude of election disinformation and platform reluctance to act.

The recommendations below are modest in their ambition not because the challenge is small but out of humility before the hard path to progress. Some of them are fine-tuned to respond to immediate problems facing individual researchers and their work; others are broad strategic adjustments that will take time to implement. A few require policy shifts that, while difficult, are probably also necessary in the long term.

116 Meta's new Threads app does not have an explicit policy against election disinformation, although company officials say other Meta policies also apply to Threads. Kerr, D. (2023, July 27). [Meta's Threads needs a policy for election disinformation, voting groups say](https://www.npr.org/2023/07/27/meta-threads-policy-election-disinformation). *NPR*. [perma.cc/B9D3-2XZW]

Short-to-Medium Term Steps to Protect Researchers & Mitigate Harm

1. Funders, research institutions, and nonprofits should create shared resources and practices for researchers under attack.

Opponents of election integrity seek to suffocate the field by discrediting its practitioners, weighing them down with legal baggage, and denying them funding and partnership opportunities. This attack on the collective field demands a collective response. Funders, researchers, and other stakeholders should coordinate to create shared pools of resources that affected individuals can use for security, legal, communications, or other forms of support. Similar efforts exist for other fields, such as a fund for investigative reporters facing libel lawsuits, and could be replicated here.¹¹⁷ Funders should also encourage or require grantees to build security and mental health support for staff into their budgets—and back that mandate with resources.

Universities and other institutions that host disinformation researchers should make a plan in advance for dealing with sustained attacks from partisan media and political activists. Too often, communications professionals do not understand the nature of these attacks until it is too late to respond effectively. Having a crisis PR response plan already written and approved would help.

117 Farr, M. (2022, December 12). [Launching a Legal Defense Fund for Journalists](https://www.niemanreports.org/launching-a-legal-defense-fund-for-journalists). *Nieman Reports*. [perma.cc/F689-AW9R]

In isolation, counter-disinformation professionals have proven vulnerable to efforts to remove them from their positions, cast doubts on their work, and even force them to flee their homes for fear of violence—so the community should speak earlier and more loudly in support of its members.¹¹⁸ Professionals in this field should also coordinate to share mutual support and best practices in the face of political attacks.

2. Counter-election-disinformation initiatives should pivot to year-round harm reduction strategies like pre-bunking, training for election officials, and advocacy efforts.

Efforts to provide individual examples of disinformation, threats, and other unmoderated content to platforms in previous elections have fallen short, and the challenges to that model have only deepened. Many independent counter-election-disinformation initiatives are already pivoting to pre-bunking (preemptive counter-messages about disinformation themes), coordination with election officials, and other means of response. They should continue doing so.

On some level, these approaches represent an acknowledgment that systematic limits on the spread of election falsehoods are probably a long way off, if they are possible at all. Until such limits are realized, independent initiatives should prioritize engaging directly with the targets of harm, such as voters who might be misled, communities who might be disenfranchised, and election officials who might be the target of violence.

Because narratives about voting and election fraud circulate year-round and continue to influence legislation and other policies, initiatives should put additional emphasis on their activities outside of election season. Surges in support around election season

118 Adler, W.T. & Maréchal, N. (2023, August 21). [To Protect Elections, Protect Researchers](https://perma.cc/DD94-KWQX). *Center for Democracy & Technology*. [perma.cc/DD94-KWQX]

allow researchers to set up monitoring and rapid response efforts, but more consistent support could contribute to stronger, more persistent infrastructure for advocacy and counter-messaging around threats to U.S. elections.

3. Advocates should focus less on content and more on mitigating the impact of disinformation “superspreaders.”

Researchers and advocates who continue to focus on reforming platforms should reduce efforts to improve piece-by-piece content moderation and focus on the largest distributors and amplifiers of election untruths.¹¹⁹ This approach is based on actors and behavior, not content.¹²⁰ These actors might be influential individual accounts, or they may be web domains or the social media presence of organizations. Advocacy and research efforts should increase pressure for platforms to act against “superspreaders.”¹²¹

One step would be to advocate for changes to platform design and content moderation policies designed to prevent influential accounts from amplifying viral falsehoods. As Glenn Ellingson said in his interview, many content moderation remedies are “utterly pointless” when content has already spread far and wide, and so “mitigation not removal should be the response.” Such mitigation might include more platform efforts to determine sources of consistently misleading or unsafe content and penalize its distribution. In a Carter Center report titled “The Big Lie and Big Tech,”¹²² Michael Baldassaro, Katie Harbath, and Michael Scholtens

119 For example: Bond, S. (2021, May 14). [Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows.](#) *NPR*. [perma.cc/YYZ5-N28B]

120 François, C. (2019, September 20). [Actors, Behaviors, Content: A Disinformation ABC. Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses.](#) *Transatlantic Working Group*. [perma.cc/F432-HJWR]

121 Aspen Digital. (2021, November 15). [Aspen Institute’s Commission on Information Disorder Makes Recommendations to Address the Mis- and Disinformation Crisis](#) [Press release]. [perma.cc/6VJA-HGEW]

122 Baldassaro, M., Harbath, K., & Scholtens, M. (2021, August). [The Big Lie and Big Tech: Misinformation Repeat Offenders and Social Media in the 2020 U.S. Election.](#) *The Carter Center*. [perma.cc/SMS8-6RCV]

suggest various ways of doing this, such as limiting direct shares of that content, applying labels warning viewers that content comes from sources that repeatedly share false claims, and removing such sources from recommendation features and search results.¹²³

Medium-to-Long Term Strategic Shifts for Election Integrity & Advocacy

4. Researchers, donors, and advocates should treat election disinformation as part of a larger, institutional problem by supporting reforms to electoral process and law.

The United States is in the midst of perhaps the most widespread, ferocious assault on representative government since the Jim Crow era. Election disinformation in 2020 seeded the ground for new restrictions on voting rights ahead of 2022 and 2024. Counter-

123 In some cases, platforms may also ban accounts for egregious, repeated spread of content which violates terms of service. While that is not our recommendation here, it is worth mentioning a sophisticated debate about the tradeoffs of this approach. Consider two recent research findings: first, deplatforming is the most effective way to reduce the audience of repeat spreaders of mis- and disinformation; but used too liberally, deplatforming can drive that audience and more ordinary banned users onto other, less moderated platforms where they become more extreme. These findings suggest that it may be important to limit the reach of some of the internet's most dangerous information sources, but that deplatforming can sometimes backfire. See: Buntain, C., Innes, M., Mitts, T., & Shapiro, J. (2023, March 12). [Cross-Platform Reactions to the Post-January 6 Deplatforming](https://perma.cc/3Y9P-GTG4). *Journal of Quantitative Description: Digital Media*. [perma.cc/3Y9P-GTG4]; Rauchfleisch, A. & Kaiser, J. (2021, June 15). [Deplatforming the Far-right: An Analysis of YouTube and BitChute](https://perma.cc/M4LR-PTML). *SSRN*. [perma.cc/M4LR-PTML]; Ali, S. Saeed, M.H., Aldreabi, E., Blackburn, J., De Cristofaro, E., Zannettou, S., & Stringhini, G. (2021, June). [Understanding the Effect of Deplatforming on Social Networks](https://perma.cc/WS83-NQ9M). *Proceedings of the 13th ACM Web Science Conference 2021*. [perma.cc/WS83-NQ9M]

disinformation as a broader field should consider the wider set of incentives and institutions that need reform in order to make U.S. elections more fair and representative.

In the long term, protecting elections from disinformation cannot rest on accountability for big tech alone. Many researchers recognize that technology is not solely responsible for disinformation; it is a reflection of deeper societal ills which will ultimately require combining solutions within the technology sector with solutions beyond it.

An example is the first-past-the-post primary system used by most states to nominate political candidates. Coupled with gerrymandered, uncompetitive districts, this process essentially makes nominating contests a race to the extremes: winning the base is tantamount to winning the general election. As a result, candidates have incentives to use irresponsible rhetoric and cater

**In the long term,
protecting elections
from disinformation
cannot rest on
accountability for big
tech alone.**

to extreme demands, often veering into outright disinformation. Secretaries of State withdrawing from election pacts like ERIC are prime examples,¹²⁴ as are election deniers who triumph in primaries.

Contributing factors like these might be addressed through reforms such as ranked choice voting, which can reward moderate candidates with broad appeal; reforms to redistricting processes to promote competitiveness and fairness; reforms to election

advertising laws to limit the use of generative AI and personal data for ad targeting¹²⁵; improvements to post-election auditing procedures and communication¹²⁶; and enhancing election officials' ability to have a strong and trusted web presence.¹²⁷ Already

124 Parks, M. (2023, June 6). How the far right tore apart one of the best tools to fight voter fraud. *NPR*. [perma.cc/42XL-QPF9]

125 Consider: Kleinfeld, R. (2022, September 15). Five Strategies to Support U.S. Democracy. *Carnegie Endowment for International Peace*. [perma.cc/LAB4-MPEX]

126 Adler, W.T. (2022, October 31). De-Weaponizing and Standardizing the Post-Election Audit. *Center for Democracy & Technology*. [perma.cc/Z26A-GE4B]

127 Adler, W.T., Doyle, J., Kiran, M.M., Jones, M.L., & Ohm, P. (2022, October 19). Only 1 in 4 Election Websites Uses the .gov Domain. That's a Problem — and an Opportunity. *Center for Democracy & Technology*. [perma.cc/NH46-UTHQ]

ongoing are bipartisan efforts to offer more support and protection to election officials.¹²⁸ More attention could also be paid to the detrimental role of cable news television and the decline of local, more civically healthy media.

5. Advocates and their donors should increase the resources spent on advocacy in select states.

Tech regulation in Congress has moved in a slow, seemingly endless grind, but many state legislatures have proven more agile. Federal regulation provides the benefits of standardization and coverage: it preempts the problem of compliance across competing jurisdictions and provides rights and benefits to every U.S. resident. But improvements in states can serve as models and, in some cases, platforms may react by adapting consistent practices across the country.

New laws on election advertising, data privacy, security for election workers and polling places, researcher access to data, and other relevant issues could be promoted in key states. California would be a priority due to its high population and the fact that it is home to several major social media companies; swing states like Michigan, Arizona, and Pennsylvania would also be logical places to start. Rather than lead this charge themselves, national funders and nonprofits should find local policy advocates and partner with them on key issues.

128 Rodriguez, E. & Patton, M. (2023, July 6). One quiet bipartisan way state legislatures are making election administration stronger. *Protect Democracy*. [perma.cc/4MFD-A3F4]; Fernekes, C., Harbath, K., & Buck, M.B. (2022, August 24). How Tech and Election Officials Can Protect Elections Online. *Bipartisan Policy Center*. [perma.cc/68G4-WCSX]

6. Platform Improvements to Capacity, Process, Oversight, & Accountability.

Platforms should re-commit to moderation of election disinformation, reinvest in trust & safety, and explore non-regulatory forms of industry accountability and oversight.

Platforms should re-commit to addressing election disinformation, reverse recent pullbacks in their policies, and reinvest in trust and safety teams—especially those specializing in election and civic integrity issues. They should also continue exploring voluntary and co-regulatory approaches to accountability and oversight.

The decision to shed trust and safety jobs despite critical threats to election integrity in the United States following the January 6th insurrection was irresponsible. Platforms should reverse course as soon as possible and begin restaffing trust and safety teams, focusing especially on civil rights subject matter experts and empowering them to influence platform policies and content moderation practices.

In order to counter claims of censorship and bias, platforms should adhere to widely accepted principles such as the Santa Clara Principles on Transparency and Accountability in Content Moderation. These principles were developed by leading digital rights organizations and academic experts and have been endorsed by major platforms; their full implementation would lead to content moderation that is less arbitrary, more rights-respecting, and more transparent.

In the United States, the First Amendment poses formidable obstacles to, and in many cases forecloses, government efforts to regulate platform content moderation directly. Voluntary and co-regulatory schemes present an alternative route. Such approaches have been pursued in Europe—for example, the 2022 EU Code of Practice on Disinformation was designed with input from both platforms and regulators and encourages several positive steps such as more detailed, more uniform, country-level sharing of

data on disinformation.¹²⁹ This Code of Practice is likely to become formalized under the Digital Services Act Article 45 as a Code of Conduct, which will guide the Commission in evaluating and enforcing DSA obligations. These approaches benefit from the involvement of civil society—something advocates criticized the original EU Code of Practice for lacking.¹³⁰

Independent, non-government organizations can also play a role in fostering accountability. In 2019, Meta (then Facebook) created the Oversight Board, an independent body of human rights and free expression experts who rule on content moderation appeals and issue advisory opinions to the platform. The Oversight Board provides a layer of accountability where there previously was not one. Even if that layer has holes, it also had a tangible impact: the Board has ruled on questions as consequential as the suspension of former President Trump’s account¹³¹ and increased public knowledge of Meta’s COVID-19 policies¹³² and its “cross-check” program for reviewing moderation decisions on highly visible accounts.¹³³

7. Platforms should voluntarily disclose more about their communications with government agencies.

Platforms should increase transparency about their communications with governments to help restore public confidence in counter-election-disinformation efforts. Platforms do not have to wait for the government to implement their own transparency reforms; they can

129 European Commission. (n.d.). [The 2022 Code of Practice on Disinformation](#). [perma.cc/QL4E-HEQM]

130 Consider the following statements: Access Now, AlgorithmWatch, Civil Liberties Union for Europe, and European Digital Rights. (2022, February 24). [Joint Statement on Stakeholder Inclusion in the Code of Practice on Disinformation Revision Process](#). [perma.cc/EA4W-RNP3]; EU DisinfoLab. (2022, September 8). [Position of the EU DisinfoLab on the 2022 Code of Practice on Disinformation](#). [perma.cc/9MF8-67V5]

131 Oversight Board. (n.d.). [Former President Trump’s suspension](#). [perma.cc/77AA-87E8]

132 Oversight Board. (n.d.). [Removal of COVID-19 misinformation](#). [perma.cc/R39M-5JS8]

133 Oversight Board. (n.d.). [Meta’s cross-check program](#). [perma.cc/5HWR-DQF4]

disclose content-related conversations on their own, similar to the way that they have released information on government requests for access to personal data.¹³⁴ An example might be the Lumen Database maintained by the Berkman Klein Center for Internet and Society at Harvard, which compiles requests provided by platforms and other sources for content takedowns for copyright and other reasons.¹³⁵

8. Platforms should create consistent points of contact for civil society.

The personalized nature of communications between civil society and researchers on the one side and platforms on the other contributes to platform unresponsiveness. This was a challenge even before layoffs on trust and safety teams, and the layoffs have only worsened the issue. Platforms should designate teams as well as individuals as appropriate points of contact, institutionalize relationships so they do not wither when key staff depart and empower designees to escalate outside concerns within companies.

9. Platforms should expand researcher access to platform data.

Researcher access to platform data is a common recommendation for counter-disinformation.¹³⁶ But in the United States, platforms have begun to regress in this area despite regulatory efforts to compel data sharing in Europe.

134 Meta. (n.d.). [Government Requests for User Data](https://perma.cc/5FZ2-N5QW). [perma.cc/5FZ2-N5QW]

135 Lumen. (n.d.). [About Us](https://perma.cc/BFR6-YEU7). [perma.cc/BFR6-YEU7]

136 For example: Pasquetto, I.V. et al. (2020, December 9). [Tackling misinformation: What researchers could do with social media data](https://perma.cc/7XXH-2A88). *Harvard Kennedy School Misinformation Review*. [perma.cc/7XXH-2A88]

The public deserves to understand how today's most important communication tools are impacting politics and society. The corporations responsible for these tools should not wait for lawmakers to act; instead, they should voluntarily create responsible, robust processes for data-sharing with independent researchers.

In interviews, some platform staff were wary of calls for access to back-end data for outside researchers. They fear the possibility of abuses like those during the Cambridge Analytica scandal and the risk that platforms that share more will receive disproportionate public criticism. These are not strong arguments against data-sharing; rather, they illustrate the importance of a strong governance regime that mitigates the risk of abuse.

Government Steps to Promote Trust & Safety and Public Confidence in Elections

10. Government and independent institutions should promote and make use of former trust & safety staffers' talent.

The Digital Forensic Research Lab (DFRLab) report on "Scaling Trust on the Web" notes that trust and safety is "shifting from a community of practice into a field." It highlights organizations like the Trust & Safety Professionals Association and the Integrity Institute as important professional bodies for an expert community of professionals.¹³⁷

137 DFRLab. (2023, June 21). *Scaling Trust on the Web: Comprehensive Report of the Task Force for a Trustworthy Future Web*. [perma.cc/D7T3-JD44]

Layoffs across the tech sector forced many members of this community out of the corporate world. Many of them are passionate about this work and have found new jobs in advocacy or research organizations, where they continue to promote trust & safety from the outside. As Jesse Littlewood from Common Cause said, these individuals can be an asset and more should be done to build bridges between them and advocacy organizations. Many nonprofits have already done well in hiring these professionals.

Funders and government agencies should support this trend by systematically supporting the development of an independent trust and safety field. They can do so by helping it establish a greater number of institutionalized centers for knowledge and exchange outside of the corporate sector, where they can promote accountability. The DFRLab report wisely suggests that trust & safety learn from cybersecurity and other adjacent fields which have grown to include significant training opportunities, diversified hiring pipelines, and government guidelines to cultivate best practices.

11. Governments should clarify and be more transparent about their role in responding to election disinformation.

Though it has since been narrowed, the July 4th, 2023, court injunction in *Missouri v. Biden* demonstrates increasing political risks from government engagement with platforms and researchers on election disinformation. If elements of the injunction go into full effect, the breadth of its exceptions and continuing court battles around them ensure that future rules on this engagement will require clarification.

To increase public insight into these relationships and ward off future political conspiracism, Congress—or, barring that, the executive branch on its own—should establish formal transparency processes to disclose when agencies communicate with platforms, the substance of those communications, any content specifically flagged, and any actions taken as a result. These processes will have to take into account security and privacy concerns (for example by anonymizing sample content) while still allowing for meaningful oversight.

State election authorities do not have to wait for action at the federal level. State legislatures should, where possible, lay out their own rules governing transparency about engagement with social media platforms and disinformation researchers.

Methodical Consideration of Generative AI and its Potential Risks

12. Researchers and the government should carefully assess the potential impact of generative AI and prioritize responses based on risk.

It is still too early to judge the impact of generative AI on disinformation as a political challenge. Further, efforts to regulate generative AI and disinformation may quickly run into First Amendment issues.

Researchers and the government should conduct a thorough review of the possible risk that generative AI will increase the problems around disinformation, beginning by acknowledging questions to which the answers are not yet known. What are the most likely forms of harm experts can anticipate, and what are the best ways to mitigate or prevent them? Is there evidence that AI-generated content is more persuasive than other forms of political speech, and to whom, in what contexts? How are bad actors likely to try and disseminate this content?

Some potential steps for addressing AI-generated disinformation are already percolating. For example, some leading AI companies have voluntarily committed to the White House to “[d]evelop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated, including robust provenance,

watermarking, or both, for AI-generated audio or visual content.”¹³⁸ The efficacy of these efforts is yet to be determined, but setting norms and expectations could help mitigate some harm while creating incentives for better labeling systems down the line.

Certain forms of consumer education will also almost certainly be required. As generative AI spreads, more bespoke models will become available; it is already no longer the sole province of large corporations.¹³⁹ But a personally trained chatbot devoid of guardrails against misinformation is more like a mirror than a manipulator. As with disinformation, the technological problem is only an aspect of the larger social challenge. Stakeholders should avoid a narrow focus on technical solutions and instead focus on how this technology is most likely to be used by individuals and the likely impact of that use. This could open up non-technical avenues for response in areas like political advertising, consumer protection, and civil and criminal law.

138 OpenAI. (2023, July 21). [Moving AI governance forward](https://perma.cc/9QTT-ZNRN). [perma.cc/9QTT-ZNRN]

139 Thompson, S.A. (2023, July 8). [Uncensored Chatbots Provoke a Fracas Over Free Speech](https://perma.cc/76D3-NZTD). *New York Times*. [perma.cc/76D3-NZTD]

06

Conclusion

In March 2023, internet scholar Kate Klonick wrote a counterintuitive essay entitled “The End of the Golden Age of Tech Accountability” in which she argues that “2021 was a heyday for trust and safety,” a time when tech companies felt public pressure to take a number of positive (if insufficient) self-regulatory steps.¹⁴⁰ She laments that platforms are now backtracking as a result of economic headwinds and the failure of many governments to pass meaningful regulation while public outrage was at its peak. A few months later, in June 2023, the prominent technology journalist Casey Newton cited Klonick’s argument in a newsletter, asking, “Have we reached peak trust and safety?”¹⁴¹

140 Klonick, K. (2023, March 3). [The End of the Golden Age of Tech Accountability](https://perma.cc/F9B6-EMD8). *The Klonickles*. [perma.cc/F9B6-EMD8]

141 Newton, C. (2023, June 8). [Have we reached peak trust and safety?](https://perma.cc/A7DE-YEV8) *Platformer*. [perma.cc/A7DE-YEV8]

The trends detailed in this report will probably tempt most readers to answer “yes.” There are many reasons to be pessimistic about prospects for improvement. But improvement is possible if the field accepts that election disinformation is an environmental hazard to be managed, not a disease to be cured. Few signs in the near term point to huge gains in the health of the U.S. media ecosystem. Steps can be taken to protect and better support researchers, diminish the prevalence and severity of harm, achieve incremental improvements in tech accountability and transparency, and set up the trust and safety field for long-term success.



07

Appendix: List of Interviews

This report draws from 29 interviews with 31 individuals. Below is a list representing interview participants; most interviews are listed by organization to protect the identities of individuals who asked to remain anonymous. In some cases, we interviewed more than one individual affiliated with an organization; in others, multiple individuals joined the same interview. Some individuals were interviewed based on a previous affiliation and may be included under it here. In cases where even naming a professional affiliation might reveal a participant's identity, we have instead characterized their relationship to this work in broader terms.


Interviewees


- Anti-Defamation League
- Brennan Center for Justice
- Center for Internet Security
- Common Cause (Jesse Littlewood and others)
- Digital Forensic Research Lab
- Disinfo Defense League (Staff and Member Organizations)
- Election Integrity Partnership
- Freedom House
- Google
- Institute for Strategic Dialogue
- João Guilherme Bastos Dos Santos (Brazilian Scholar, interviewed in no affiliated capacity)
- Katie Harbath (AnchorChange)
- Leadership Conference for Civil and Human Rights
- Meta
- Microsoft
- National Democratic Institute
- Nina Jankowicz (Center for Information Resilience)
- Oversight Board
- Paula Gori and Nikos Sarris (European Digital Media Observatory)
- Researcher subpoenaed by the House Judiciary Committee
- The Carter Center
- The Integrity Institute
- Twitter




 cdt.org

 cdt.org/contact

 **Center for Democracy & Technology**
1401 K Street NW, Suite 200
Washington, D.C. 20005

 202-637-9800

 @CenDemTech

