

The Data Driving Democracy

Understanding How the Internet Is
Transforming Politics and
Civic Engagement



Christina Couch

AMERICAN ACADEMY OF ARTS & SCIENCES

The Data Driving Democracy

Understanding How the Internet Is
Transforming Politics and
Civic Engagement

Christina Couch

© 2020 by the American Academy of Arts & Sciences.
All rights reserved.

ISBN: 0-87724-130-9

This publication is available online at www.amacad.org/project/practice-democratic-citizenship.

Suggested citation: Christina Couch, *The Data Driving Democracy: Understanding How the Internet Is Transforming Politics and Civic Engagement* (Cambridge, Mass.: American Academy of Arts and Sciences, 2020).

Cover image: © iStock.com/400tmax.

This paper is part of the American Academy's Commission on the Practice of Democratic Citizenship. The statements made and views expressed in this publication are those held by the author and do not necessarily represent the views of the Officers and Members of the American Academy of Arts & Sciences.

Please direct inquiries to:
American Academy of Arts & Sciences
136 Irving Street
Cambridge, MA 02138
Telephone: 617-576-5000
Fax: 617-576-5050
Email: aaas@amacad.org
Web: www.amacad.org

Contents

v	Summary
1	Introduction and Context
3	Report Goals and Methodology
5	Data for Digital Age Democracy
11	Key Insights: What Can These Data Tell Us?
27	Barriers and Challenges
27	Data Access
29	Ethical Challenges
31	Research Infrastructure Insufficiencies
34	Potential Solutions
34	Data Sharing Initiatives
37	Regulation and Legal Solutions
40	Acknowledgments
42	About the Author

Summary

This report outlines the data and methodologies researchers use to understand how the Internet has impacted democracy and the challenges they face in this field. The report summarizes key insights from interviews with fifteen experts from a broad array of computer science, data analysis, media studies, legal, and political science backgrounds. It specifically examines the data and research methodologies experts use to study how the Internet is changing democracy, the types of inferences that can (and can't) be drawn with current resources, and barriers in this field.

Key Insights:

- Current data resources, tools, and methodologies can sufficiently track how specific messages spread across individual platforms and channels, who they target, and how they change over time, but understanding how messages ripple through the larger media ecosystem across multiple channels is an open question. Understanding their impact on beliefs and behaviors is also an open question.
- Current data resources, tools, and methodologies can offer valuable insights into specific demographic pockets in ways that were not possible before the digital age, but these online data are not representative samples of the general population. The insights gained come with limitations on how far they can be extrapolated.
- Platforms can be crucial tools for galvanizing grassroots social movements and elevating marginalized voices, but these spaces and tools largely reinforce existing power structures and biases along race, gender, and socioeconomic lines. Researchers are limited in the ways in which they can interrogate how those power dynamics are established and their ramifications.
- Search engines and social media platforms play an increasingly powerful role in political speech, voter knowledge, and democratic participation, but there is not enough transparency around curatorial and ranking algorithms nor on how policies within tech companies are crafted and executed.
- Despite the challenges, many experts believe in the Internet's potential to promote democracy and strengthen civic engagement, though everyone who spoke on this topic said that online spaces

and the tech sector at large have a long way to go to achieve these positive results.

When asked about what barriers hinder their research, experts pointed to data access challenges, legal concerns, ethical issues, and insufficient academic research infrastructures. They discussed data sharing initiatives, regulation, and lawsuits as possible solutions to some of these problems.

Introduction and Context

It started with “a love letter to Black people.”¹ In the summer of 2013, when George Zimmerman was found not guilty of second-degree murder in the death of seventeen-year-old Trayvon Martin, Alicia Garza, then an activist and community organizer in Oakland, California, took her feelings to Facebook. She wrote an impassioned post: “I continue to be surprised at how little Black lives matter. And I will continue that. stop giving up on black life.” She ended with the following words: “black people. I love you. I love us. Our lives matter.” Patrisse Cullors, a friend and fellow activist, added a hashtag and announced that #BlackLivesMatter was a movement. Opal Tometi, an immigration rights activist in Brooklyn, built the group’s social media strategy.

#BlackLivesMatter gained traction as an organizing tool and online campaign for amplifying anti-racism voices.² In 2014, when eighteen-year-old Michael Brown was killed by a White police officer, Black Lives Matter (BLM) became a mobilizing force behind nationwide protests. It has since grown to encompass a broad spectrum of work from anti-racism and Black empowerment groups, and has given rise to the Black Lives Matter organization,³ the Campaign Zero policy reform group,⁴ a political action committee (PAC), and at least one BLM Super PAC.⁵

Born and fueled by digital political activism, #BlackLivesMatter showcased the Internet’s capacity to mobilize social and protest movements, quickly disseminate political messages, and amplify new voices. Frank Leon Roberts, an activist who teaches a BLM course at New York University,

1. Jelani Cobb, “The Matter of Black Lives,” *The New Yorker*, March 6, 2016, <https://www.newyorker.com/magazine/2016/03/14/where-is-black-lives-matter-headed>.
2. Patrisse Khan-Cullors, “We Didn’t Start a Movement. We Started a Network,” *Medium*, February 22, 2016, <https://medium.com/@patrissemariacullorsbrignac/we-didn-t-start-a-movement-we-started-a-network-90f9b5717668>.
3. Black Lives Matter, <https://blacklivesmatter.com>.
4. Campaign Zero, <https://www.joincampaignzero.org>.
5. Lateshia Beachum, “A Super PAC Has Raised Millions to Mobilize Black Voters. Does It Matter That Its Funders are White?” *The Center for Public Integrity*, March 12, 2018, <https://publicintegrity.org/politics/a-super-pac-has-raised-millions-to-mobilize-black-voters-does-it-matter-that-its-funders-are-white>; and “Statement of Organization,” Black Lives Matter PAC LLC, September 22, 2015, <https://docquery.fec.gov/pdf/304/20150922030023304/201509220300023304.pdf#navpanes=0>.

called it “the first U.S. social movement in history to successfully use the Internet as a mass mobilization device.”⁶

For some, BLM and other cultural moments galvanized online, including #MeToo, #TimesUp, and #NoBanNoWall, exemplify ways that the Internet can upend civic engagement, creating a mode of political participation tailor-made for the rising distrust in traditional legislative and media institutions and redistributing power to underrepresented citizens. The Internet has also worked in the opposite direction by bolstering existing political power structures, offering new ways to manipulate behavior and information flow, catalyzing hate campaigns, and creating an opaque landscape where users aren’t aware of what biases they are being subjected to or how their data are being used.

How the Internet has transformed the practice of democracy and how the on- and offline worlds of political participation bleed into each other are questions that become more complicated to answer every second that goes by. These questions encompass everything from non-state election interference and the spread of disinformation to voting security and political polarization to petition and crowdfunding avenues and mainstream media manipulation. They reach from the highest levels of governments worldwide all the way down to daily small talk in communities across the geographic and socioeconomic spectrums.

The rise of online civic engagement presents a veritable mountain of new data, communities, tools, and ways to study political participation as well as new ways to augment and reframe traditional research in this sphere. This report explores what data and methodologies researchers use to understand the Internet’s impact on the democratic process and the obstacles that prevent them from getting the answers they seek. For the purposes of this report, civic engagement is broadly defined and includes voting, political volunteer work, fundraising and individual-level donations, protest activity, attending political meetings, boycotting and politically motivated consumer spending, on- and offline engagement with representatives, political mobilization, broadcasting political messages and opinions, and working within organizations to influence the political landscape.

6. Frank Leon Roberts, “How Black Lives Matter Changed the Way Americans Fight for Freedom,” ACLU, July 13, 2018, <https://www.aclu.org/blog/racial-justice/race-and-criminal-justice/how-black-lives-matter-changed-way-americans-fight>.

Report Goals and Methodology

The goal of this report is to identify the key indicators experts believe point to how this field is transforming and to showcase how researchers track and study online civic engagement. Specifically, this report examines:

- The data and research methodologies those within the field use to explore how the Internet is changing democracy.
- The types of inferences they believe can (and can't) be reliably drawn using these resources.
- The barriers that prevent researchers from understanding this field better.

The influence of online activity and platforms on civic engagement is an enormous topic, one that already fills several books. This report is not comprehensive. It does not lay out all the questions facing those who work in the online civic engagement space, nor does it systematically address all data sources, research methods, and research or researchers operating within this field. It is also not intended to be an objective piece of journalism. It is instead designed to give a snapshot to a lay reader of work from a handful of leading researchers—fifteen to be exact—and to provide a broad overview of common practices, takeaways, and challenges from their perspectives. Though the author hopes that this report offers something of interest for researchers who work in this field, the report was written for a more general audience who may not be familiar with the underlying data and methodologies that drive this area of study.

Experts included in this report were selected to represent a wide spectrum of work happening within this field. They include computer scientists, data analysts, political scientists, sociologists, media studies scholars, legal experts, and others, but the list of individuals interviewed is by no means exhaustive. Many important voices are not found within the pages of this report, mostly due to time and resource constraints. The report summarizes key points from these interviews and incorporates additional research for context, but it does not provide a comprehensive view of the field.

One major limitation of this report is that it does not include insights from those working on the industry side. This was a conscious choice made

in part because of length constraints and because industry perspectives continuously receive widespread media attention. Titans like Facebook and Twitter have their own platforms to broadcast their messages as well as access to mainstream media resources that aren't as readily available to individual researchers. It is worth noting that researchers within digital platforms and tech companies have made extremely valuable contributions in this space.

Another limitation is geographic scope. This report is almost entirely focused on work happening within the United States and work that is centered around U.S.-based media, technology, and politics. As such, it does not address important social issues, valuable projects, and many major voices leading international initiatives in this space nor does it address the pivotal role the Internet played in events like the Arab Spring and the Rohingya refugee crisis. It also does not reflect the fact that the vast majority of users of social media platforms are *not* in the United States or that millions of Americans are active on social media platforms that are not based in the United States.

One final disclosure: This report highlights one expert and some research from the Massachusetts Institute of Technology (MIT). The author also works for MIT, but not in a public relations capacity.

Data for Digital Age Democracy

An article published in July of 2019 by academics from the University of California, San Diego, and the University of Massachusetts, Amherst, referred to the large quantity of new data available to leaders in the civic engagement space as the “Civic Data Deluge.”⁷ It’s an imperfect term that captures the vast amount of “big data” that is now available to researchers, but leaves out the tsunami of information that isn’t accessible because of technical, legal, and proprietary ownership barriers.

To measure how civic engagement plays out on- and offline, researchers use an enormous array of data resources, each of which come with their own content limitations and, in many cases, accessibility challenges. Diving into each individual data resource would be exhausting to read, not to mention impossible as these resources are constantly changing. This section goes over briefly the types of data that drive research in this arena and how academics obtain these data.

Much of the research in this field, especially studies around political messaging, fake news and disinformation, online campaign reach, and media manipulation, centers around social media data from widely popular channels like Facebook, Twitter, and YouTube, and to a lesser extent community-specific channels like Gab,⁸ a free speech-focused social media platform known for attracting racists and far-right extremists. Experts interviewed for this report noted that social media is particularly valuable to civic engagement research because it offers immediate ways to observe how online movements catalyze and mobilize,⁹ measure the reach of

7. Narges Mahyar et al., “The Civic Data Deluge: Understanding the Challenges of Analyzing Large-Scale Community Input,” *Proceedings of the 2019 Designing Interactive Systems Conference*, June 23–28, 2019, 1171–1181, http://groups.cs.umass.edu/nmahyar/wp-content/uploads/sites/8/2019/05/DIS_2019_The_Civic_Data_Deluge_Final-4.pdf.

8. Yuchen Zhou et al., “Elites and Foreign Actors Among the Alt-Right: The Gab Social Media Platform,” *First Monday* 24 (9) (September 2, 2019), <https://firstmonday.org/ojs/index.php/fm/article/view/10062/8072>.

9. Deen Freelon, Charlton D. McIlwain, and Meredith D. Clark, “Beyond the Hashtags,” The Center for Media & Social Impact, American University, February 2016, https://cmsimpact.org/wp-content/uploads/2016/03/beyond_the_hashtags_2016.pdf.

media outlets,¹⁰ draw inferences based on how users casually talk about politics,¹¹ and follow how political messages move and morph throughout the digital ecosphere.¹²

Henry E. Brady, dean of the Goldman School of Public Policy at the University of California, Berkeley, gave an example of how social media can streamline research.¹³ For two of his books, Brady and his coauthors tracked messaging from a broad spectrum of political lobbying groups.

“In the past, if you wanted to study that, what you would have had to have done is somehow figure out what the lobbying groups were, subscribe as a member, and then maybe get the information they sent you via mail and then content analyze that,” he explained, later adding that using this method, it was not possible to single out messages from individual lobbyists working within an organization. “Instead, we just got all the Twitter feeds from these organizations and then we content analyzed them using data science methods and came to some conclusions about what was going on. I mean, that’s an enormously useful thing to know.”

Social media data often include data from the offline world as well. Facebook, for example, has location tracking capabilities that are on even when users aren’t using the app.¹⁴ Google came under fire in 2018 for storing location data from Android and iPhone devices even when users opted out of sharing location information.¹⁵ Instagram offers shopping checkout that gathers data on purchases made through partner merchants.¹⁶ Experts described the breadth of digital trace data gathered by social media and search engine platforms as “enormous” and “mind-boggling,” noting that

10. Yochai Benkler et al., “Study: Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda,” *Columbia Journalism Review* (March 3, 2017), <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>.

11. Pablo Barberá et al., “Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?” *Psychological Science* 26 (10) (August 21, 2015): 1531–1542, <https://journals.sagepub.com/doi/abs/10.1177/0956797615594620?journalCode=pssa>.

12. David M. J. Lazer et al., “The Science of Fake News,” *Science* 359 (6380) (March 9, 2018): 1094–1096.

13. Interview conducted with Henry E. Brady by phone on August 5, 2019.

14. Paul McDonald, “Understanding Updates to Your Device’s Location Settings,” Facebook, September 9, 2019, <https://about.fb.com/news/2019/09/understanding-updates-to-your-devices-location-settings/>.

15. Ryan Nakashima, “AP Exclusive: Google Tracks Your Movements, Like it or Not,” Associated Press, August 13, 2018, <https://apnews.com/828aefab64d4411bac257a07c1af0ecb>.

16. Instagram Business Team, “New to Instagram Shopping: Checkout,” Instagram, March 19, 2019.

it includes information on users' physical activities, consumer behaviors, and information consumption.

The exact data that researchers can access from social media vary dramatically from platform to platform, which is a primary reason why these types of civic data studies heavily skew toward more public and accessible platforms and away from more restrictive ones. Academics typically obtain social media data in one of two ways: 1) through an API (Application Programming Interface), which is a tool companies offer that allows third parties to access a limited set of data curated by the company, and 2) through scraping, a process wherein extraction tools are used to go through sites and pull unstructured data specified by the researcher.

Both methods, alone and in combination with collection and sentiment analysis tools (a partial list of about one hundred of these tools is available in a report by Lily Davies¹⁷), are incredibly powerful and both come with challenges. APIs are generally designed for third-party developers rather than academics and they range in terms of the types of data that can be accessed. Since APIs are designed and controlled by tech companies, they can (and do) change,¹⁸ which can disrupt ongoing research. APIs also frequently come with hefty fees and "rate limits"¹⁹ that restrict the number of data requests parties can make within a certain timeframe. Data scraping provides an alternative route to data access. This method is often technically challenging and time-consuming for researchers, but the primary barrier to this method of access is legality. Many platforms prohibit automated data scraping²⁰ and include language in their terms of service that severely limits how researchers can use public data that have been pulled from the site without automated software. In some cases, researchers can also purchase some consumer spending data through third-party aggregation firms; however, in these instances, researchers do not have access to the raw data and must work with information that has already been manipulated by an outside entity. A further discussion of these issues is in the Barriers and Challenges section of this report.

17. Lily Davies, "Social Media Data in Research: A Review of the Current Landscape," Sage Ocean, June 4, 2019, <https://ocean.sagepub.com/blog/social-media-data-in-research-a-review-of-the-current-landscape>.

18. Mike Schroepfer, "An Update on Our Plans to Restrict Data Access on Facebook," Facebook.com, April 4, 2018.

19. "Rate Limiting," Twitter.com, accessed September 2, 2019, <https://developer.twitter.com/en/docs/basics/rate-limiting>.

20. "Prohibited Software and Extensions," LinkedIn.com, accessed September 2, 2019, <https://www.linkedin.com/help/linkedin/answer/56347/prohibited-software-and-extensions?lang=en>.

Studies that revolve around the Internet's impact on democracy also mine data from myriad other online networks, including search engines, comment boards, wikis, blogs, chat apps, location-based programs, mainstream media and government sites, online petition sites, and large datasets provided by government agencies, nonprofits, nongovernmental organizations, and private companies. Though some organizations like Google²¹ and Democracy Works²² offer their own civic information-specific APIs, many of the same access challenges and limitations apply to these resources as well. Experts said that data structure also plays a role here. Because platforms vary in terms of what data are available and how they are organized, researchers have an incentive to focus on some platforms over others.

"If you look at Twitter, tweets have a nice structure to them and there are certain kinds of metadata associated with a tweet, like when was it tweeted and who tweeted it. A tweet is a fairly neat object to work with and that scales up when you have lots of tweets," said David Lazer, co-director of Northeastern University's NULab for Texts, Maps, and Networks and professor of political science and computer and information science.²³ "Reddit has threads and there are lots of reddit [channels] so it's just a more complicated data structure. Reddit is this sort of very threaded thing that's much tougher to work with."

The field isn't entirely reliant on digital data and quantitative analysis. Journalistic research, like Kate Klonick's analysis of how platforms develop procedural systems for governing online speech,²⁴ and ethnographic research, such as Francesca Tripodi's report on media messaging and interpretation within conservative Christian communities,²⁵ offer nuanced answers to questions that quantitative methods can't fully address. Some work, such as Jen Schradie's research on class divides within digital political

21. "Google Civic Information API," [developers.google.com](https://developers.google.com/civic-information/?hl=en), accessed September 3, 2019, <https://developers.google.com/civic-information/?hl=en>.

22. "Democracy Works Elections API," [Democracy.works](https://www.democracy.works/elections-api/), accessed September 3, 2019, <https://www.democracy.works/elections-api/>.

23. Interview conducted with David Lazer by phone on August 8, 2019.

24. Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech," *Harvard Law Review* 131 (6) (April 10, 2018): 1598–1670, https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf.

25. Francesca Tripodi, "Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices," *Data & Society*, May 16, 2018, https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Searching-for-Alternative-Facts.pdf.

participation,²⁶ uses a mixed methods approach. Researchers noted that there are fewer qualitative than quantitative studies in this field, in part because of the time and cost of conducting good qualitative studies and partially due to the general trend in social science research toward quantitative analysis. Many experts interviewed for this report spoke of a need for more qualitative research in this field. This report focuses primarily on research that relies on quantitative data, and thus the many fine qualitative studies that do exist are not discussed as much as they might be in a longer report.

Both quantitative and qualitative projects often use digital data in tandem with resources relied upon before the digital age, such as voter registration data, phone surveys, polls, public records, judicial opinions, campaign contribution data, census research, and so on. The Internet has also transformed how these traditional data resources are used and what they mean. For example, telephone public opinion polls that had response rates of 37 percent in the late 1990s dropped to response rates of just 6 percent as of 2018.²⁷ Low response rates translate to higher survey costs and reduced sample sizes as well as a bias toward older demographics. For these reasons, major polling organizations such as the Pew Research Center²⁸ now conduct the majority of U.S. polling online, but data from online surveys aren't entirely comparable to phone polling data.²⁹ In opinion polls that involve speaking with a live interviewer, respondents are significantly more likely to give answers that sidestep awkward interactions and that frame themselves and their communities in positive ways, which skews research centered around socially undesirable attitudes and behaviors.³⁰

26. Jen Schradie, "The Digital Activism Gap: How Class and Costs Shape Online Collective Action," *Social Problems* 65 (1) (February 2018): 51–74, <https://academic.oup.com/socpro/article/65/1/51/4795348>.

27. Courtney Kennedy and Hannah Hartig, "Response Rates in Telephone Surveys Have Resumed Their Decline," Pew Research Center, February 27, 2019, <https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/>.

28. Courtney Kennedy and Claudia Deane, "What Our Transition to Online Polling Means for Decades of Phone Survey Trends," Pew Research Center, February 27, 2019, <https://www.pewresearch.org/fact-tank/2019/02/27/what-our-transition-to-online-polling-means-for-decades-of-phone-survey-trends/>.

29. Courtney Kennedy, "Phone vs. Online Surveys: Why Do Respondents' Answers Sometimes Differ by Mode?" Pew Research Center, February 7, 2019, <https://www.pewresearch.org/fact-tank/2019/02/07/phone-vs-online-surveys-why-do-respondents-answers-sometimes-differ-by-mode/>.

30. Scott Keeter et al., "From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls," Pew Research Center, May 13, 2015, <https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/>.

Respondents also answer complex questions differently when they hear the question being read to them versus reading it themselves.³¹ This is to say that the work of this field involves not only understanding what the available digital data can reveal about how citizens engage with politics, but also how new modes of political participation have forced research methodologies to evolve. The reliance on research models that existed before the rise of the Internet presents another challenge: that of understanding what it is that we want to measure in the first place. A focus on getting access to data, whether from phone surveys or from online social media platforms, privileges the questions that these sorts of data can answer. Interrogating the foundations of how we think about civic and political participation is necessary if researchers are to better understand the health of our civic life.

31. Jon A. Krosnick and Duane F. Alwin, "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement," *Public Opinion Quarterly* (1987), <https://academic.oup.com/poq/article-abstract/51/2/201/1847197>.

Key Insights: What Can These Data Tell Us?

Interviews for this report included one extraordinarily broad and borderline unfair question: *From your perspective, what questions can the available data and research methodologies answer, and what questions can't be answered in this field?* This section details a handful of key insights that highlight how experts view their field of study. The insights showcased here do not comprehensively cover digital political participation research, but are instead intended to give readers an introduction to what the landscape of online civic engagement research looks like to those on the inside.

1. Current data resources, tools, and methodologies can efficiently track how specific messages spread across individual platforms and channels, who they target, and how they change over time, but understanding how messages ripple through the larger media ecosystem is still an open question. Understanding their impact on beliefs and behaviors is also an open question.

In 2012, a local Florida news station aired a small story about the shooting of an unarmed teenager named Trayvon Martin. The *Orlando Sentinel* and the *Miami Herald* newspapers also ran articles on Martin's death, but the story nearly stopped there. It wasn't until ten days later that the national news media broke the story to a broader audience. The shooting was brought to national attention largely thanks to Benjamin Crump, a civil rights attorney who took on Martin's case pro bono, and to publicist Ryan Julison. Martin's story didn't just spread from mainstream news media to race-based media outlets, activist sites, and a Change.org petition backed by celebrity voices; it pivoted along the way, from a story framed as an altercation between two people to one centered around an unarmed Black teenager dying at the hands of a neighborhood watch vigilante who wasn't held accountable.

Readers can follow the exact path that Martin's story took to get from that first Fox 35 Orlando piece to national protests and remarks from the

President³² as well as how the story changed along the way thanks to media mapping research from Erhardt Graeff, Matt Stempeck, and Ethan Zuckerman.³³ To map Martin's story, the team used quantitative data from eight different sources—RSS feeds collected with the Media Cloud media analysis tool, front page national newspaper stories, broadcast television news mentions, Google searches for “Trayvon Martin,” Google searches for “George Zimmerman,” tweets, Change.org petition signatures, and clicks on bitly links within Media Cloud stories—combined with firsthand interviews.

Ethan Zuckerman, director of the Center for Civic Media and associate professor of the practice in media arts and sciences at the Massachusetts Institute of Technology, was the principal investigator on the paper. He recounts how the project was perceived by others in the field: “Someone cited Erhardt's Trayvon Martin paper and basically said, ‘The researchers did this using eight different data sources. Obviously, that is insane,’” Zuckerman said, laughing. “My response was I get it, but it's not insane. Actually we felt bad that we didn't get certain other data sources into that study.”³⁴

Media ecosystem studies that follow messages as they move between platforms and throughout the larger digital universe are far more rare than studies that track how messages spread on one specific platform. Experts largely agreed that current data and research methodologies are effective at identifying influential voices, issues users are talking about, accounts that are exhibiting problematic behaviors, and how messages spread within a specific platform. Many experts pointed to the entire subfield of research on the mechanisms by which fake news and disinformation spread—we have several examples³⁵—as crucial additions to digital civic engagement literature and as proof that valuable, useful conclusions can be drawn from single platform studies.

32. Barack Obama, “Remarks by the President on Trayvon Martin,” The White House, July 19, 2013, <https://obamawhitehouse.archives.gov/the-press-office/2013/07/19/remarks-president-trayvon-martin>.

33. Erhardt Graeff et al., “The Battle for ‘Trayvon Martin’: Mapping a Media Controversy Online and Off-line,” *First Monday* 19 (2) (February 4, 2014), <https://journals.uic.edu/ojs/index.php/fm/article/view/4947/3821>.

34. Interview conducted with Ethan Zuckerman by phone on August 22, 2019.

35. Soroush Vosough et al., “The Spread of True and False News Online,” *Science* 359 (6380) (March 9, 2018): 1146–1151, <https://science.sciencemag.org/content/359/6380/1146>; Kate Starbird et al., “Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter After the 2013 Boston Marathon Bombing,” *iConference 2014 Proceedings*, 2014, https://faculty.washington.edu/kstarbi/Starbird_iConference2014-final.pdf; and Andrew Guess et al., “Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook,” *Science Advances* 5 (1) (January 9, 2019), <https://advances.sciencemag.org/content/5/1/eaau4586>.

“You can answer lots of questions that look at how different segments of society use specific social media platforms,” said Deen Freelon, associate professor in the Hussman School of Journalism and Media at the University of North Carolina at Chapel Hill.³⁶ “You can get a sense of how information circulates within the platforms. There’s some information about the effects that it has, who ends up engaging with it, how information flows between the people that originate it if they’re not already media professionals who are well known, and how it originates, how it flows towards folks that have more visibility and reach. Those are the kinds of questions that I think are well-answered.”

Experts also agreed that single platform studies alone can’t provide a comprehensive view of how influence moves and gets amplified throughout the Internet nor whether these messages actually impact beliefs and behaviors. Since both political outreach and disinformation campaigns are often designed to push users to further engagement across platforms—think Twitter posts that link to YouTube videos that link to blogs—single platform studies only offer a small piece of a much larger picture.

“As great as it is to have papers that are about how a certain thing travels on Facebook or on Twitter with regards to Myanmar or Mexican elections or Indian elections, none of us only exist in that one space and there are other spaces that are more important in different places in the world,” said Alondra Nelson, president of the Social Science Research Council and Harold F. Linder Chair in the School of Social Science at the Institute for Advanced Study.³⁷ “It’s all well and good to pay a research team to do some research on the polarization situation on Twitter vis-à-vis a certain issue, but how are we going to understand how and when that moves from WhatsApp to Instagram? How different generations are using different apps to [stoke] virality? The role of bots? It’s such a complicated thing.”

Experts were quick to point to what they viewed as seminal ecosystems research, including Yochai Benkler, Robert Faris, and Hal Roberts’ 2018 book *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*,³⁸ which presents a map of the American political and media landscape during the 2016 presidential election (a 2017 study³⁹

36. Interview conducted with Deen Freelon by phone on August 23, 2019.

37. Interview conducted with Alondra Nelson by phone on September 10, 2019.

38. Yochai Benkler et al., *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford: Oxford University Press, 2018), <https://www.oxfordscholarship.com/view/10.1093/oso/9780190923624.001.0001/oso-9780190923624>.

39. Rob Faris et al., “Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election,” The Berkman Klein Center for Internet & Society, Harvard University, August 16, 2017, <https://cyber.harvard.edu/publications/2017/08/mediacloud>.

by the same authors plus several additional ones preceded the book), and Zeynep Tufekci's 2017 book *Twitter and Tear Gas: The Ecstatic, Fragile Politics of Networked Protest in the 21st Century*,⁴⁰ which examines the role of the Internet in modern protest movements. They were also quick to note the barriers that prevent researchers from doing more multi-platform ecosystem studies, which range from a lack of technical tools to make data collection and analysis easier across platforms to research infrastructures that inhibit interdisciplinary work to lack of coordination between stakeholders within academia and the tech industry.

An ecosystems approach “might mean that we need to have journalists working with Facebook and have [platforms offer] clear understandings of the technologies for journalists,” said Joan Donovan, director of the Technology and Social Change Research Project at Harvard Kennedy School's Shorenstein Center on Media, Politics and Public Policy.⁴¹ “We need to have civic society organizations be able to report to platform companies quickly that there's something happening within their communities that is suspect or that there is some kind of manipulation campaign. We also need university researchers to be able to access data and to be able to audit platforms so that their research isn't so patchwork.” Donovan clarified that the data she is referring to are related to platforms' revenue, advertising, and manipulation campaigns; not data about individual users.

2. Current data resources, tools, and methodologies can offer valuable insights into specific demographic pockets in ways that were not possible before the digital age, but these online data are not representative samples of the general population. Insights gained come with limitations on how far they can be extrapolated.

When discussing how the Internet has changed the way researchers measure civic engagement, many experts brought up demographic granularity. Existing data resources and tools provide new ways (and more streamlined ways) to identify and study highly specific demographics and subpopulations as well as smaller, community-specific websites, blogs, social media, and other communication networks. Just as digital platforms and “big data” resources have given rise to microtargeted political ads and outreach strategies, they've also created new mechanisms for answering detailed, community-specific questions like who speaks up about housing development at planning and zoning board meetings in

40. Zeynep Tufekci, *Twitter and Tear Gas: The Ecstatic, Fragile Politics of Networked Protest in the 21st Century* (New Haven: Yale University Press, 2017), <https://www.twitterandteargas.org/>.

41. Interview conducted with Joan Donovan by phone on August 27, 2019.

eastern Massachusetts⁴² and how do local governments leverage social media in crises.⁴³

Experts noted that digital data also provide new ways for researchers to study political attitudes and ideologies that are commonly considered socially sensitive or outright unacceptable. One example a few experts cited is Seth Stephens-Davidowitz's study that looked at whether the percentage of "racially charged" Google searches made in specific geographic areas during the 2008 and 2012 elections could predict Barack Obama's vote share in those places, controlling for vote shares of the previous Democratic presidential candidate.⁴⁴ (Spoiler: They did.) In the case of that particular study, Google searches provided a way to collect aggregate data from a large number of people, a method of pegging search terms to geographic locations, and a window into racial perceptions and ideologies that are often difficult to study using traditional survey methods.

Many experts issued a word of warning with regard to studies that analyze language, keywords, and political speech: Current quantitative methods often have difficulty factoring in tone or context, even when using sentiment analysis tools. That means that it's often difficult using quantitative methods alone to figure out the intention behind some messages and whether they're true expressions of how a user actually feels. This obstacle is especially limiting when tracking words like "climate" or "immigration" that have political and nonpolitical connotations, and when studying communities that have their own language and slang conventions. Getting the lingo wrong muddies research results. This points to another challenge of researching political and civic participation online across a range of platforms that are designed to present text alongside video alongside photographs. The variety of content types presented online makes conducting comprehensive research extremely difficult. Researchers' efforts have been based primarily on text-based online material; we do not yet have the tools to work meaningfully with audio, video, and image content, which represents a large percentage of the content consumed online.

42. Katherine Levine Einstein et al., "Who Participates in Local Government: Evidence from Meeting Minutes," *Perspectives on Politics* 17 (1) (March 2019): 28–46, <https://www.cambridge.org/core/journals/perspectives-on-politics/article/who-participates-in-local-government-evidence-from-meeting-minutes/C6505940E607B6392C4A8F53A9363DB1>.

43. Melissa W. Graham et al., "The Role of Social Media in Local Government Crisis Communications," *Public Relations Review* 41 (3) (September 2015): 386–394, <https://www.sciencedirect.com/science/article/abs/pii/S0363811115000077>.

44. Seth Stephens-Davidowitz, "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data," *Journal of Public Economics* 118 (October 2014): 26–40, <https://www.sciencedirect.com/science/article/abs/pii/S0047272714000929>.

Jonathan Nagler, co-director of the New York University Social Media and Political Participation (SMaPP) Lab and director of the NYU Politics Data Center, said that machine learning text analysis programs are only as good as the datasets they're trained on.⁴⁵ Creating high-quality datasets—ones that account for nuances in language—require careful design, human coders, and quality control measures, all of which require time and financial investment. Even with ample fiscal and human investment, research methodologies will still struggle to keep up with speech ambiguities and with breaking stories and events that introduce new buzzwords and phrases into the political vernacular. “The constraint is going to be can we as a social science community have the resources to build quality training datasets for the task at hand,” Nagler said.⁴⁶

While online data and data tools have made it easier to pinpoint some specific demographics and how some individual messages travel within platforms, experts reported that they still struggle to find digital data samples that are representative of the general population. That's partially because despite having large numbers of users, platforms, search engines, and datasets come with biases that make it difficult to know if inferences drawn from those data can be extrapolated to the broader public.

About 70 percent of Americans use social media,⁴⁷ but the use is wildly uneven. Twitter, for example, skews more heavily toward younger, liberal, and higher income users. Nearly one-quarter of all Americans use the platform, but just 10 percent of Twitter users create 80 percent of the site's content⁴⁸ and the platform is flooded with content-producing bots that sound increasingly more like real people.⁴⁹ That means that tools like Twitter's Sample Tweets API,⁵⁰ which offers a free random sample of

45. Interview conducted with Jonathan Nagler by phone on August 21, 2019.

46. A further discussion of the methodological challenges of creating quality machine learning datasets for political text analysis is available in a paper co-authored by Nagler. See Pablo Barberá et al., “Methodological Challenges in Estimating Tone: Application to News Coverage of the U.S. Economy,” presented at the annual meeting of the American Political Science Association, Philadelphia, Pennsylvania, September 2016, https://as.nyu.edu/content/dam/nyu-as/faculty/documents/econmedia_mpsa2016_methods_rvBBLMN.pdf.

47. “Social Media Fact Sheet,” Pew Research Center, June 12, 2019, <https://www.pewresearch.org/internet/fact-sheet/social-media/>.

48. Stefan Wojcik and Adam Hughes, “Sizing Up Twitter Users,” Pew Research Center, April 24, 2019, <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>.

49. Luca Luceri et al., “Evolution of Bot and Human Behavior During Elections,” *First Monday* 24 (9) (September 2, 2019), <https://firstmonday.org/ojs/index.php/fm/article/view/10213/8073>.

50. “Snapshot of Tweets in Real-Time,” Twitter, accessed September 10, 2019, <https://developer.twitter.com/en/products/tweets/sample>.

all public tweets in real time, pulls the majority of content from a small, disproportionately vocal population.⁵¹ YouTube and Facebook are, by far, the most widely used platforms—both are heavily used across every age group except those over age sixty-five. In some cases, platforms also skew along racial and ethnic lines. Instagram and WhatsApp, for example, are much more popular among Hispanic users compared to other ethnic demographics.⁵²

Biases inherently mean that some demographics are left out, oftentimes low-income groups and those low in socioeconomic status (more on that in the next section).⁵³ Additionally, researchers are not able to study what's said privately on these platforms. Even if they have access to all public data, that's still only a slice of all platform activity and, as one expert said, it's often "not the most interesting slice." Perhaps due to privacy concerns, many Americans are now retreating to closed communication forums that offer higher levels of encryption. An unknown number of Americans also use social media platforms that are based outside the United States, several of which (China's WeChat is one example) seamlessly integrate multiple functions into a single platform.

"I worry that we're missing large numbers of people," Henry Brady said. "When we go and get a whole lot of Facebook posts, we just simply don't know what that's representative of . . . it's worrisome that we don't actually have a good notion of what universe they represent. That's, I think, the biggest problem with Internet data—you just don't know what the universe is."

Experts said that data access problems further complicate this issue by preventing researchers from clearly understanding the biases and limitations of their samples and by inhibiting research replication. Experts also noted that this sampling problem is one that political and social science grappled with in various forms long before the digital age. However, they added, big data digital resources create the illusion of representation and a certain objectivity that's impervious to human biases—a phenomenon Microsoft principal researcher Kate Crawford calls "data

51. A further discussion of problems with Twitter's sampling tools can be found in Jürgen Pfeffer et al., "Tampering with Twitter's Sample API," *EPJ Data Science* 7 (1) (December 19, 2018), <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0178-0>.

52. Andrew Perrin and Monica Anderson, "Share of U.S. Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018," Pew Research Center, April 10, 2019, <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>.

53. Kay Lehman Schlozman et al., *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy* (Princeton: Princeton University Press, 2012).

fundamentalism”⁵⁴—while data access and other problems prevent researchers from fully understanding exactly who and what exactly they’re studying.

3. Platforms can be crucial tools for galvanizing grassroots social movements and elevating marginalized voices, but these spaces and tools largely reinforce existing power structures and biases along race, gender, and socioeconomic lines. Researchers are limited in the ways in which they can interrogate how those power dynamics are established and their ramifications.

Jen Schradie, a sociologist at the Observatoire Sociologique du Changement (OSC) at Sciences Po in Paris, is part of a pool of researchers whose work shows that the problems of representation inherent with digital civic engagement data and exacerbated by platform design often translate to the powerful becoming more powerful. In the late 2000s, Schradie was intrigued by the Internet’s promise to democratize politics. Online, anyone could have a political voice, but Schradie wondered who was actually producing the content that would drive these voices.

Using survey data from roughly forty-one thousand American adults, Schradie analyzed ten ways of creating digital content, ranging from chat-room participation to blog production, and found that, just like in traditional media, digital content production varied along socioeconomic lines. Across all ten production activities, users who had higher levels of formal education were more likely to be content producers. They also produced significantly more content than those lower on the ladder,⁵⁵ indicating that instead of disrupting existing power structures, the Internet had strengthened voices that were already louder than others.

Schradie and a handful of other researchers have repeatedly shown that online civic engagement spaces, platforms, and tools have evolved, but the digital divide along class and other lines remains. Schradie’s subsequent work has broadly shown class and education gaps in who produces blogs⁵⁶

54. Kate Crawford, “The Hidden Biases in Big Data,” *Harvard Business Review*, April 1, 2013, <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

55. Jen Schradie, “The Digital Production Gap: The Digital Divide and Web 2.0 Collide,” *Poetics* 39 (2) (April 2011): 145–168, <https://www.sciencedirect.com/science/article/abs/pii/S0304422X1100012X>.

56. Jen Schradie, “The Trend of Class, Race, and Ethnicity in Social Media Inequality,” *Information, Communication & Society* 15 (4) (March 13, 2012): 555–571, <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.665939>.

and levels of digital activism within grassroots groups.⁵⁷ In one study, she conducted in-depth interviews and ethnographic research, and examined social media posts from thirty-four activist groups all organized around the issue of collective bargaining and unionization rights for public employees in North Carolina. Out of sixty thousand total tweets posted within the study's timeframe, all but one had come from middle- and upper-class groups.⁵⁸

This skew toward wealthier users is largely because content production and digital messaging take time, labor, and resources, all of which tend to be far more available to those higher on the socioeconomic chain, Schradie said. It's also worth noting that the classic measures of civic and political participation have been rooted in structures of power and privilege that shape decisions about what to measure in the first place. Other experts noted that marginalization is further enforced by bots, social media manipulation, and computational propaganda campaigns that send armies of automated cybertroops to amplify certain messages thousands of times per day and quash others, as well as by algorithms that incentivize sensationalism. Researchers added that these automated programs can outmatch marginalized voices in terms of the volume of posts they make, but it is not clear how much messages from bots drive (or don't drive) behavior.

Researchers are still struggling to quantify the extent of influence that bots have over online political discourse and offline behavior, but there's no doubt that bots are extensive, prolific, and sometimes dominant voices around politically charged conversations, in some cases driving more than half of social media conversations around specific topics.⁵⁹ Platforms have taken steps over the past few years to crack down on malevolent bots,⁶⁰ and some research suggests that certain steps may help to curtail the problem.

57. Jen Schradie, *The Revolution That Wasn't: How Digital Activism Favors Conservatives* (Cambridge, Mass.: Harvard University Press, 2019).

58. Jen Schradie, "The Digital Activism Gap: How Class and Costs Shape Online Collective Action," *Social Problems* 65 (1) (February 2018): 51–74, <https://academic.oup.com/socpro/article/65/1/51/4795348>.

59. Issie Lapowsky, "Here's How Much Bots Drive Conversation During News Events," *Wired*, October 30, 2018, <https://www.wired.com/story/new-tool-shows-how-bots-drive-conversation-for-news-events/>; and Sergey Sanovich, "Computational Propaganda in Russia: The Origins of Digital Misinformation," Computational Propaganda Working Paper Series No. 2017.3, Oxford Internet Institute, 2017, <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Russia.pdf>.

60. Yoel Roth, "Automation and the Use of Multiple Accounts," Twitter, February 21, 2018, https://blog.twitter.com/developer/en_us/topics/tips/2018/automation-and-the-use-of-multiple-accounts.html.

Earlier this year, for example, WhatsApp put harsher restrictions on the number of times a user can forward a specific message,⁶¹ reducing the limit from a maximum of twenty groups down to five. In a preprint paper published on Arxiv.org, researchers from Universidade Federal de Minas Gerais in Brazil and the Massachusetts Institute of Technology analyzed public data gathered from WhatsApp in Brazil, India, and Indonesia and found that the new limits didn't eradicate propaganda, but they did slow the spread of misinformation by about one order of magnitude.⁶²

Despite the crackdowns, studies show that organized disinformation campaigns are still growing. The Computational Propaganda Research Project at the Oxford Internet Institute documented social media manipulation campaigns in seventy countries in 2019—that's up from forty-eight the year before—with Facebook being the favored platform among bad actors in fifty-six nations.⁶³ Experts interviewed for this report credited Alice Marwick and Rebecca Lewis's 2017 Data & Society report on media manipulation campaigns as a crucial text for understanding how disinformation propagates.⁶⁴

"It really comes down to money, power, and resources," says Schradie.⁶⁵ "Whether you are an individual who is just kind of creating bots on your own, you have to have time to do that. More often, it's an organization, an institution, or a state government."

Schradie's conclusions that digital participation is dominated by the elite and that online spaces and platforms more often than not bolster those who already have political power is echoed by a body of scientific

61. Fanny Potkin et al., "Facebook's WhatsApp Limits Users to Five Text Forwards to Curb Rumors," Reuters, January 21, 2019, <https://www.reuters.com/article/us-facebook-whatsapp/facebook-whatsapp-limits-text-forwards-to-five-recipients-to-curb-rumors-idUSKCN1PF0TP>.

62. Philippe de Freitas Melo et al., "Can WhatsApp Counter Misinformation by Limiting Message Forwarding?" *Arxiv.org*, September 23, 2019, <https://arxiv.org/pdf/1909.08740.pdf>.

63. Samantha Bradshaw and Phillip N. Howard, "The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation," Computational Propaganda Research Project, 2019, <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>.

64. Alice Marwick and Rebecca Lewis, "Media Manipulation and Disinformation Online," Data & Society, May 15, 2017, http://www.chinhnghia.com/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf.

65. Interview conducted with Jen Schradie by phone on August 27, 2019.

literature and activist work from across the globe.⁶⁶ Experts who were involved in this area of research pointed to a broad array of projects that document this phenomenon— from Zeynep Tufecki’s work on the ways in which social media makes protest groups vulnerable⁶⁷ to Cass Sunstein’s research on echo chambers and political polarization⁶⁸ to David Karpf’s projects around analytic activism⁶⁹ to Cathy O’Neil’s work chronicling how big data reinforce discrimination along race, sex, and economic lines while appearing neutral.⁷⁰

Experts also gave a nod to the breadth of research centered around how the broader tech sector unintentionally reinforces racial, gender, and economic inequality, both in and outside of political spheres. Work like Safiya Umoja Noble’s research that documents how search engine algorithms reinforce racism⁷¹ and Ruha Benjamin’s work on discriminatory design⁷² are directly applicable to questions around how the Internet influences civic engagement.

Experts were also quick to say that data access problems and algorithmic opacity prevent researchers from understanding the full extent of these dynamics and from creating solutions (more on these issues in the next section). Many said that they were disturbed and concerned by what we know about biases and political power structures, but they were more concerned about the hidden issues that can’t be uncovered yet because the tools to do so don’t exist and because the data (if they exist at all) are not available.

66. See Tufecki, *Twitter and Tear Gas*; Cass R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton: Princeton University Press, 2017); and Maggie Dwyer and Thomas Molony, eds., *Social Media and Politics in Africa: Democracy, Censorship and Security* (London: Zed Books, 2019).

67. Tufecki, *Twitter and Tear Gas*.

68. Sunstein, *#Republic*.

69. David Karpf, *Analytic Activism: Digital Listening and the New Political Strategy* (Oxford: Oxford University Press, 2016).

70. Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing Group, 2016).

71. Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

72. Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity, 2019).

4. Search engines and social media platforms play an increasingly powerful role in political speech, voter knowledge, and democratic participation, but there is not enough transparency around curatorial and ranking algorithms nor on how policies within tech companies are crafted and executed.

There's clear evidence on how heavily social media, search engines, and online communities influence some civic engagement metrics like galvanization of protest and social movements,⁷³ spread of political messages,⁷⁴ voter turnout,⁷⁵ and ability to reach younger voters.⁷⁶ There's conflicting evidence about how much these platforms influence other metrics like political perceptions⁷⁷ and whether political messages translate to changes in behavior or belief (a question that has plagued the study of media forms long before the rise of the Internet).

Experts agreed that social media and search engines, in particular, are becoming increasingly significant to how political messages spread from both the politician and citizen sides, but the exact ways that tech companies influence what users see are unclear and users themselves are frequently unaware. Research shows that users often don't know that search

73. Deen Freelon, Charlton D. McIlwain, and Meredith D. Clark, "Beyond the Hashtags," The Center for Media & Social Impact, American University, February 2016, https://cmsimpact.org/wp-content/uploads/2016/03/beyond_the_hashtags_2016.pdf; and Tufecki, *Twitter and Tear Gas*.

74. Andrew Chadwick and Cristian Vaccari, "News Sharing on UK Social Media: Misinformation, Disinformation, and Correction," Online Civic Culture Centre, Loughborough University, May 2, 2019; and Alexandre Bovet and Hernán A. Makse, "Influence of Fake News in Twitter During the 2016 U.S. Presidential Election," *Nature Communications*, January 2, 2019, <https://www.nature.com/articles/s41467-018-07761-2>.

75. Katherine Haenschen, "Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout," *Journal of Communication* 66 (4) (2016): 542–563, <https://onlinelibrary.wiley.com/doi/abs/10.1111/jcom.12236>; and Robert M. Bond et al., "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, September 12, 2012, <https://www.nature.com/articles/nature11421>.

76. "Five Takeaways on Social Media and the Youth Vote in 2018," The Center for Information and Research on Civic Learning and Engagement, Tufts University, November 15, 2018, <https://civicyouth.org/five-takeaways-on-social-media-and-the-youth-vote-in-2018/>.

77. R. Kelly Garrett, "Social Media's Contribution to Political Misperceptions in U.S. Presidential Elections," *PLoS ONE*, March 27, 2019; and Robert Epstein et al., "Suppressing the Search Engine Manipulation Effect," *Proceedings of the ACM on Human-Computer Interaction* 1 (2) (November 2017), <https://cbw.sh/static/pdf/epstein-2017-pacmhci.pdf>.

engine results and social media feeds are curated at all.⁷⁸ Experts reported that this curatorial opacity prevents researchers from better understanding the landscape and effects of political messaging and misinformation.

Aaron Smith, director of Data Labs at the Pew Research Center, said that current research methodologies can allow researchers to follow how many users are sharing or viewing messages in the aggregate within most platforms, but algorithmic opacity makes it hard to understand the backdrop and context within which individuals encounter those messages.⁷⁹

On most major social media platforms, “you can’t just look at the people they’re following and know [with certainty] that they encountered a particular tweet or engaged with a particular type of content,” Smith said. “Drawing links between an individual person and the actual content that they’re seeing and being exposed to and engaging with on digital platforms that they use and how that bleeds into things like knowledge of elections or support of candidates or support for conspiracy theories, any question of choice, that linkage is very difficult. . . . That’s kind of the Holy Grail for what we’re trying to figure out.”

Experts said that uncovering those levels of exposure is especially relevant when tracing harmful behaviors like hate speech, harassment, and extremism, especially on networks that monetarily incentivize virality. A few people interviewed for this report praised Rebecca Lewis’s 2018 report, “Alternative Influence: Broadcasting the Reactionary Right on YouTube,”⁸⁰ as a crucial text for understanding how audiences move from mainstream to extremist content and how extremist messages are perpetuated. They added that algorithmic transparency from platforms like YouTube could greatly enhance these types of projects and provide a clearer picture of the choices users have when selecting which content to engage with. Experts said that algorithmic opacity also prevents researchers from uncovering ways that algorithms reinforce biases, especially along gender and racial lines, and it gives rise to accusations of censorship.

“Just having that transparency from the corporate level would help appease a lot of the conspiracy theories that I think are thriving right now,”

78. Motahhare Eslami et al., “I always assumed that I wasn’t really that close to [her]: Reasoning about Invisible Algorithms in News Feeds,” *Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems*, April 18, 2015, <http://www.kevinhamilton.org/share/papers/p153-eslami.pdf>; and Paul Hitlin and Lee Rainie, “Facebook Algorithms and Personal Data,” Pew Research Center, January 16, 2019, <https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/>.

79. Interview conducted with Aaron Smith by phone on September 6, 2019.

80. Rebecca Lewis, “Alternative Influence: Broadcasting the Reactionary Right on YouTube,” *Data & Society*, September 18, 2018, https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf.

said Francesca Tripodi, assistant professor of sociology at James Madison University and affiliated researcher with the Data & Society Research Institute, who studies media manipulation.⁸¹ This opacity is primarily in place to protect intellectual property and corporate financial interests, Tripodi added, but it's also because curatorial algorithms often aren't understood even by those creating them.

A few experts pointed to Safiya Umoja Noble's book, *Algorithms of Oppression: How Search Engines Reinforce Racism*,⁸² which details the myriad ways that search engines reinforce privilege and discriminate against people of color, especially women of color, as an example of work that addresses how search biases influence information consumption. While many tech companies have tweaked existing policies to increase transparency—for example, last year Facebook modified procedures for informing page managers when content was removed⁸³ and published information on how news feeds are personalized⁸⁴—experts believed that these efforts did little to meaningfully illuminate automated processes and remove barriers for study.

Some experts mentioned that better protocols for algorithmic auditing could help researchers in this field find and eliminate bias issues. While some data scientists like Cathy O'Neil have created their own algorithmic auditing consultancy groups, either independently or within established organizations, others have pushed for “right to explanation” provisions similar to those embedded in the European Union's General Data Protection Regulation (GDPR) that would require those deploying decision-making algorithms to provide an explanation to affected users of why the algorithm made the choice it did.⁸⁵

In addition to calling for greater algorithmic transparency, experts also wanted greater transparency around how tech companies create and enforce content and speech policies. Kate Klonick, assistant professor of law at St. John's University and affiliate fellow at the Information Society Project at Yale Law School and New America, authored one of the first analyses on how platforms moderate online speech through policy and the procedural

81. Interview conducted with Francesca Tripodi by phone on August 23, 2019.

82. Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

83. “Making Pages More Transparent and Accountable,” Facebook, January 23, 2019, <https://about.fb.com/news/2019/01/making-pages-more-transparent/>.

84. Ramya Sethuraman, “Why Am I Seeing This? We Have an Answer for You,” Facebook, March 31, 2019, <https://about.fb.com/news/2019/03/why-am-i-seeing-this/>.

85. Rachel Courtland, “Bias Detectives: The Researchers Striving to Make Algorithms Fair,” *Nature*, June 20, 2018, <https://www.nature.com/articles/d41586-018-05469-3>.

systems they use to develop those policies.⁸⁶ In an interview, she described the lack of knowledge around free speech and comment moderation policies as “kind of like if we didn’t have the story of the Constitution and the Founding Fathers and the American Revolution and were asked to just understand and buy into a system that surrounded us without having any idea how it got put into place.”⁸⁷

Klonick added that the need for transparency around speech policies is especially important as platforms continue to apply U.S.-centric policies to an ever-widening base of international users.

5. Despite the challenges, many experts believe in the Internet’s potential to promote democracy and strengthen civic engagement, though everyone who spoke on this topic said that online spaces and the tech sector at large have a long way to go to achieve these positive results.

Much of the research around how the Internet has transformed the political playing field examines how online spaces, tools, and overall designs have been leveraged intentionally and unintentionally to hinder democratic practice. Some (but definitely not all) researchers interviewed for this report also highlighted the opposite—the potential ways that online spaces could be optimized to amplify voices of positive change, disseminate accurate information quickly, debunk disinformation, and support healthy civic engagement in ways that reduce inequalities.

These experts pointed to several positive ways that online spaces have influenced civic engagement, including making political issues more accessible to younger audiences, leveraging personal networks to increase voter turnout, launching pro-democracy protest movements, expanding accessibility for those who cannot participate politically in-person, and broadening the definition of civic engagement. They also spoke extensively about what could be—specifically, how online spaces and tools could be optimized for political inclusion, help build civic-minded communities, fact check political figures, support positive collective action, promote accurate news, and increase faith in trustworthy voices and institutions. It is worth noting that no expert who spoke about the Internet’s ability to strengthen democracy said that the current landscape is adequately fulfilling this potential or that they believed that this potential would be fulfilled any time in the near future.

86. Kate Klonick, “The New Governors: The People, Rules, and Processes Governing Online Speech,” *Harvard Law Review* 131 (6) (April 10, 2018): 1598–1670, https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf.

87. Interview conducted with Kate Klonick by phone on August 26, 2019.

“We don’t really know the pro-democracy uses of social media because we haven’t tried,” said Ethan Zuckerman, who has written extensively on this topic.⁸⁸ “We don’t see a lot of thoughtful, meaningful deliberation on social media, but we haven’t optimized platforms for that.”

Zuckerman said that there are pockets of the web that either currently serve or have served as examples of spaces that are optimized to promote democracy and reason-based political discourse. He mentioned moderated sites like Parlio, a discussion platform that was dedicated to promoting civil opinion-sharing and debate, as one example of an online space designed to encourage fact-based ideological diversity. Parlio threads are still available online, but the site has not been updated in several years.

Zuckerman added that social media platforms are currently designed to connect users to people with whom they share things in common, but they could be optimized to counteract political polarization and break echo chambers by also connecting users with voices they might not otherwise encounter. There are several tools designed to do this, including the Center for Civic Media’s Gobo tool, which allows users to filter their social media feeds in several ways, one of them being diversification of political perspectives.⁸⁹

Some experts spoke of the inherent conflict between the goals of advertisement-driven platforms and spaces that serve public, rather than corporate, interests. A few experts said that creating pro-democracy platforms might require public funding, a conscious move away from for-profit business models, and stricter rules on appropriate advertising within those spaces. It would also require thinking critically about how to measure healthy civic engagement, inclusivity, and empowerment within these spaces.

88. Ethan Zuckerman, “We Make the Media—A Recent Speech at Freedom of Speech Online 2018,” [Ethanzuckerman.com](http://www.ethanzuckerman.com/blog/2018/12/09/we-make-the-media-a-recent-speech-at-freedom-of-speech-online-2018/), December 9, 2018, <http://www.ethanzuckerman.com/blog/2018/12/09/we-make-the-media-a-recent-speech-at-freedom-of-speech-online-2018/>; and Ethan Zuckerman, “Six or Seven Things Social Media Can Do For Democracy,” [Ethanzuckerman.com](http://www.ethanzuckerman.com/blog/2018/05/30/six-or-seven-things-social-media-can-do-for-democracy/), May 30, 2018, <http://www.ethanzuckerman.com/blog/2018/05/30/six-or-seven-things-social-media-can-do-for-democracy/>.

89. For more information about the Center for Civic Media’s Gobo tool, see <https://www.media.mit.edu/projects/gobo/overview/>.

Barriers and Challenges

Experts who were interviewed for this report highlighted three major obstacles that hinder digital civic engagement research: 1) data access and legal concerns, 2) ethical issues, and 3) insufficient research infrastructure. This section dives briefly into each of these problems and outlines a few proposed solutions in various stages of execution. This section is far from exhaustive. Fifteen experts were interviewed. All fifteen detailed an array of very specific concerns facing the field. This section touches on the problems that continually cropped up in the interviews, but by no means covers everything.

Data Access

Experts interviewed almost unanimously agreed that data access is one of the biggest obstacles that prevent researchers from getting a comprehensive understanding of how search engines, social media platforms, and other online spaces are changing democracy. Experts were clear and adamant: It is impossible to fully understand how activity on online platforms influences democracy without having more data from those platforms, as well as a better understanding of what data those platforms actually collect. The current research landscape is patchy at best, the experts said. David Lazer from Northeastern University's NULab for Texts, Maps, and Networks compared it to "looking up at the sky with no instrumentation. We can't see the stars because we don't have equipment to do that and that's a real shame." The lack of data access not only prevents researchers from getting a full picture; it also prevents studies from being replicated and results from being repeatedly confirmed over time.

Platforms limit data access through API restrictions, rate limiting, and anti-scraping technologies as noted earlier in this report. Experts also added that changes in algorithms, platform design, and company policies can affect data access as well. Many pointed to the myriad changes that Facebook made in the wake of a highly controversial 2014 study on mood manipulation⁹⁰ and the Cambridge Analytica scandal⁹¹ from last year as

90. Mike Schroepfer, "Research at Facebook," Facebook, October 2, 2014, <https://about.fb.com/news/2014/10/research-at-facebook/>.

91. Natasha Lomas, "How Facebook Has Reacted Since the Data Misuse Scandal Broke," *Tech Crunch*, April 10, 2018, <https://techcrunch.com/2018/04/10/how-facebook-has-reacted-since-the-data-misuse-scandal-broke/>.

examples of platform policy changes that radically disrupted data flow and interrupted ongoing research. Both of those events are further outlined in the Ethical Challenges section of this report. These types of changes that affect data access not only prevent researchers from pulling data from a specific platform, they can also render both software and research libraries that rely on specific APIs functionally inoperative.

Christian Sandvig, director of ESC: the Center for Ethics, Society, and Computing and professor of digital media at the University of Michigan, said that data can also be limited through restrictive terms of service agreements, some provisions of which have been found not to be legally enforceable.⁹²

“Researchers are really in a bind because they tend not to have any legal expertise and their university or their [institutional review board] might be conservative about it, so that they might say, ‘Oh, you have to follow all the rules,’ but the rules are just ridiculous,” he said.

One expert who did not want to be identified described some terms of service agreements as “utterly Orwellian” and said that it’s often unclear what exactly researchers are allowed to do with data even if they are publicly available. For example, researchers “are totally simpatico with the idea that we shouldn’t download a bunch of Twitter data and then put out an estimate that says, ‘Hey, we think Jane Doe is a Neo-Nazi.’ We should not do that. We totally understand that,” the expert said. “We certainly think we should be able to form our own internal estimate and then put out aggregate data that’s not identifying anyone saying, ‘Hey, we noticed that everyone who tweeted saying that they approve this policy is far-right or liberal.’ By Twitter’s own terms of service, we literally are not able to know that, which makes no sense. If we publish aggregate data, is Twitter going to cancel our accounts? We don’t know with certainty.”

Experts took a variety of paths to get around these issues and several praised independent projects that help researchers better access and analyze data within current platform restrictions. Some experts specifically cited Jason Baumgartner’s efforts to increase access to reddit data through pushshift.io and the Pushshift API as one example. Other researchers discussed creative ways that they’ve built their own makeshift APIs within legal boundaries by compiling publicly available data from resources like RSS feeds. They noted that these methods were generally more difficult and time-consuming than pulling data straight from an API. Several experts said that it’s obvious from reading emerging research in the field that some teams—none who were interviewed for this report—ignore legality entirely and simply scrape sites without regard to terms of service. One expert

92. Interview conducted with Christian Sandvig by phone on August 22, 2019.

said that the entire research field “lives in a little bit of fear” of the day that platform executives decide to crack down on academics who violate their terms of service.

Ethical Challenges

Even if data access problems vanished overnight, researchers would still be confronted with ethical challenges of working in this space. Privacy and consent are primary issues here. Experts said that for certain projects, they already face challenges with trying to ensure that datasets are free of identifying information. Deen Freelon said that for his own media monitoring research, getting public-facing information while excluding content like health, financial, and educational data points that are “prohibited for us to know by law” was tough, even when using machine learning tools, black-lists, and domain filtering to eliminate data that users have not consented to share.

Privacy and consent become even more complicated as research increasingly shows that accurate, and often invasive, inferences about a specific individual can be drawn even if that person’s data are not available. One analysis of 30.8 million tweets found that it’s possible to accurately predict what someone will post online just by analyzing social media posts from eight or nine of their contacts,⁹³ meaning that an individual’s privacy and ability to consent rely on a network of people. Some scholars feel that platforms’ gestures toward differential privacy (publicly sharing some information about a dataset in order to examine patterns of use while concealing data that would identify individuals) is presented as a technological fix that does not address the underlying concerns about data availability.

Experts also said that it’s not always clear what, and who, is considered a public versus a private entity. While political campaigns and figures create social media posts with the broader public in mind, ethics get muddier when using data from individuals who may only expect their public posts to be seen by a few hundred followers. Some in the field who were not interviewed for this report have also noted that for large-scale studies, getting informed consent is not always practical.⁹⁴

Francesca Tripodi said that open data access won’t solve the field’s ethical issues. Tripodi brought up a now infamous 2014 study conducted by

93. James P. Bagrow et al., “Information Flow Reveals Prediction Limits in Online Social Activity,” *Nature Human Behaviour* 3 (January 21, 2019): 122–128, <https://www.nature.com/articles/s41562-018-0510-5>.

94. “Cambridge Analytica Controversy Must Spur Researchers to Update Data Ethics,” *Nature*, March 27, 2018, <https://www.nature.com/articles/d41586-018-03856-4>.

researchers at Facebook that looked at how emotions are transferred between users through the platform's news feeds.⁹⁵ Dubbed "the Facebook mood experiment," the study was widely criticized by the research community because the experiment was run without informing users that their feeds were being manipulated or that they were being studied. "When researchers are afforded the opportunity to get inside the system, are ethical boundaries being followed?" Tripodi asked.

Many experts said that the need for best ethical practices for researchers is even more pressing in the wake of the Facebook–Cambridge Analytica scandal, the genesis of which was rooted in academic research. To provide a quick recap, the scandal originated when University of Cambridge professor Aleksandr Kogan collected Facebook data through a personality quiz app called *thisisyourdigitallife*, which harvested data from users who gave consent to have their data used for academic purposes and from their friends and contacts who did not give consent.⁹⁶ As was widely reported in the media, data from up to 87 million users⁹⁷ were sold to Cambridge Analytica and used to develop highly controversial profiling tools and to deploy targeted political ads, most notably to support Donald Trump's⁹⁸ and Ted Cruz's⁹⁹ respective presidential campaigns. Facebook maintained that Kogan's original data collection methods were legitimate, but transferring data to a third party violated the platform's terms of service.¹⁰⁰ Since then, Facebook has pivoted toward privacy and rolled out a number of privacy-enhancing features for users, including message encryption and

95. Adam D. I. Kramer et al., "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks," *Proceedings of the National Academy of Sciences* 111 (24) (June 17, 2014), <https://www.pnas.org/content/111/24/8788.full>.

96. Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian*, March 17, 2018, <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

97. Cecilia Kang and Sheera Frenkel, "Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users," *New York Times*, April 4, 2018, <https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html?module=inline>.

98. Matthew Rosenberg et al., "How Trump Consultants Exploited the Facebook Data of Millions," *New York Times*, March 17, 2018, <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html?module=inline>.

99. Patrick Svitek and Haley Samsel, "Ted Cruz Says Cambridge Analytica Told His Presidential Campaign its Data Use was Legal," *The Texas Tribune*, March 20, 2018, <https://www.texastribune.org/2018/03/20/ted-cruz-campaign-cambridge-analytica/>.

100. Paul Grewal, "Suspending Cambridge Analytica and SCL Group From Facebook," Facebook, March 16, 2018, <https://about.fb.com/news/2018/03/suspending-cambridge-analytica/>.

proposed time limits on how long posts are saved,¹⁰¹ and it has further restricted data access for both researchers and developers.¹⁰²

To create better ethical frameworks for the field, some experts said that they want better communication channels between the academic community and technology companies and better systems for holding organizations accountable for clear ethical violations. A few mentioned the need for academics to develop ethical standards for platform data use in research contexts. Deen Freelon said that institutional review boards could provide valuable resources for researchers trying to navigate ethically murky waters, but as of now, many aren't educated about the issues involved in social media research. If brought up to speed, these entities could "help researchers construct their studies in ways that balance ethics with the most effective methods."

Research Infrastructure Insufficiencies

Experts also said that academic research infrastructure is severely underfunded and unsophisticated compared to advertising and marketing infrastructure. Researchers in this field spoke extensively about the difficulties they face in studying subjects like media manipulation, misinformation campaigns, and surveillance advertising because research tools available to academics lag far behind analytics systems used by large organizations.

"If you're a large company, you're going to have apparatus for understanding client feedback that is so well-developed that you're going to be able to act on information almost instantaneously when it hits the market," Joan Donovan from the Shorenstein Center on Media, Politics and Public Policy said. "University researchers don't have access to tech on that scale because it costs millions to make."

Experts said that funding problems hit them from multiple angles: research teams are expensive, hardware that can support sufficient computational power is expensive, and data are often expensive. Sam Gill, vice president of Communities and Impact and senior advisor to the president at the John S. and James L. Knight Foundation, said that there also needs to be more investment in laying a strong pipeline for young talent to break into the field.¹⁰³

101. Mark Zuckerberg, "A Privacy-Focused Vision for Social Networking," Facebook, March 6, 2019, <https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/>.

102. Mark Zuckerberg, Facebook post, March 21, 2018, <https://www.facebook.com/zuck/posts/10104712037900071?pnref=story>.

103. Interview conducted with Sam Gill by phone on August 27, 2019.

“You need really good graduate research assistants and postdocs in order to do the research, and then you need them going out into the world to academic jobs to propagate the research and the methods and the advancements in the field, and then you need them going out to other sectors, policy-making and applied work,” he said. “We’re missing a lot of that.”

A few experts who work on qualitative and interdisciplinary research also expressed concern about a dearth of funding and grant-making resources for basic research in the area of digital politics and civic engagement. Jesse Baldwin-Philippi, associate professor in the Communication and Media Studies Department at Fordham University who studies political communication and campaigns, noted that there are many grants that fund research around interventions aimed at improving specific metrics of civic engagement and solving specific problems, such as the spread of misinformation, but there are fewer funding options for work that examines more fundamental questions about how political campaigns and advocacy groups operate online.¹⁰⁴

Baldwin-Philippi added that some political and social scientists who are focused on this type of basic civic engagement research, whether in or outside of online contexts, are especially concerned about funding in the wake of recent changes that the National Science Foundation (NSF) made to their Social and Economic Sciences Division. The changes, which went into effect on October 1, 2019, “repositioned” several NSF programs, including transforming the Political Science Program into two separate initiatives—one focused on funding basic research around security and preparedness and the other centered on “issues broadly related to attitudes, behavior, and institutions connected to decision-making processes, the provision of essential services, and accountability mechanisms.”¹⁰⁵ While NSF representatives have reported that they believe that the changes will ultimately increase funding for political science research, the American Political Science Association issued a statement expressing concern that the move could further limit the types of projects that NSF supports.¹⁰⁶

Funding is only part of the infrastructure problem. Experts also said that the academic research infrastructure is often set up in ways that silo

104. Interview with Jesse Baldwin-Philippi conducted by phone on September 4, 2019.

105. Arthur Lupia, “Dear Colleague Letter: 2019 Social, Behavioral, and Economic (SBE) Repositioning,” National Science Foundation, September 24, 2019, <https://www.nsf.gov/pubs/2019/nsf19089/nsf19089.jsp?org=NSF>.

106. “Announcement Concerning Changes in the Political Science Program at the National Science Foundation,” American Political Science Association, August 23, 2019, <https://politicalsciencenow.com/announcement-concerning-changes-in-the-political-science-program-at-the-national-science-foundation/>.

researchers and inhibit interdisciplinary work. A few experts said that their teams largely work in isolation within their institutions and they spoke of limited opportunities for collaboration. Some experts called for better partnerships and communication channels between academic groups and tech companies. Several experts interviewed for this report cited independent research institutes that have tech company backing as viable ways to move the field forward. The Data & Society Research Institute, which was originally supported by funding from Microsoft, was mentioned by multiple experts as an example of best partnership practices.

There are projects in the works that are aimed at solving some of these issues. Several experts pointed to Social Science One's data sharing initiative (more on that project and its challenges in the next section) and to the Knight Foundation's \$39 million investment in grants for cross-disciplinary research aimed at understanding how technology is transforming democracy¹⁰⁷ as examples. Knight Foundation funds are being distributed to eleven American research institutions and think tanks that will support the creation of five new interdisciplinary centers of study. The Knight Foundation is also supporting an additional \$11 million in research that looks at Internet and digital platform governance.¹⁰⁸ The goal, Sam Gill said, is to provide insights on pressing policy questions and to help accelerate this emerging research field for the long-term future.

Some researchers are attacking academic infrastructure issues by creating shared resources that can be used to better understand this research landscape. At the Shorenstein Center, for example, Joan Donovan's team is compiling one hundred case studies that document how misinformation travels across the web. This shared digital research infrastructure, called the Global Media Manipulation Case Book, is designed to teach those who contend with media manipulation, including researchers, policy-makers, and journalists, how to spot and debunk organized campaigns.¹⁰⁹

107. "Knight Invests \$50 Million to Develop New Field of Research Around Technology's Impact on Democracy," Knight Foundation, July 22, 2019, <https://knightfoundation.org/press/releases/knight-fifty-million-develop-new-research-technology-impact-democracy/>.

108. "Funding Opportunity: Governance, Norms, and Values—Research on the Future Internet," Knight Foundation, <https://knightfoundation.org/funding-opp-research-norms-rules-governance-internet-digital-platforms/>.

109. "Technology and Social Change (TaSC) Research Project," Harvard Kennedy School Shorenstein Center on Media, Politics, and Public Policy, <https://shorensteincenter.org/about-us/areas-of-focus/technology-social-change/>, accessed October 5, 2019.

Potential Solutions

This section provides a brief overview of the potential solutions experts mentioned. It is not intended to suggest which solutions are the best or most viable; it is only intended to touch on some activity happening in this space that experts deemed important.

Data Sharing Initiatives

Over the past eighteen months, everyone in civic engagement research has been talking about Social Science One, although the conversation has taken multiple turns. Conceived by Gary King of Harvard University and Nathaniel Persily of Stanford Law School as a model for mutually beneficial partnerships between private-sector platforms and independent researchers, Social Science One sought to organize a commission of senior academic advisors, a small number of whom were bound by confidentiality agreements. These advisors would act as trusted third-party data brokers and work with platforms to identify and organize relevant datasets for use in mutually agreed-upon research projects. First, the datasets would be verified to ensure that they're not cherrypicked by the platform, then Social Science One would solicit proposal requests from the outside research community that would be reviewed by ethics and peer review boards made up of an international group of anonymous reviewers. These anonymous reviewers would evaluate the proposals for merit, but there were a couple of catches: Winning proposals can't "violate privacy or existing legal agreements/obligations" or "put a company at a competitive disadvantage."

Gary King, director of Social Science One and the Institute for Quantitative Social Science at Harvard as well as the Albert J. Weatherhead III University Professor at Harvard, said that the commission would provide a buffer for the company, as well as a mechanism by which researchers can verify that the data they receive are real and the platform wasn't "just selecting datasets to make the company look good." After a project was approved, there would be no restrictions or approval processes regarding what researchers write within the context of their proposal.

Alondra Nelson from the Social Science Research Council added that, as of August 2019, there had been no indication that a scenario that would involve rejecting a proposal based on the grounds outlined above would be a realistic possibility.

Once selected, scholars would receive funding, gain “privacy-preserving” data access, and retain the right to publish without restrictions from the company.¹¹⁰ Grants were supported by a handful of charitable foundations, which ranged from the John S. and James L. Knight Foundation to the Charles Koch Foundation.¹¹¹

Social Science One sought to take a first step toward the challenge of solving data access issues for civic engagement researchers while still protecting privacy and corporate concerns. Facebook signed on as the first partner for a research initiative dedicated to studying how social media impacts elections and the democratic process.¹¹² The first round of Social Media and Democracy Research Grant recipients was announced in April 2019.¹¹³

But the project has hit major obstacles. As of August 2019, Facebook had not yet made the promised proprietary data available, citing the inability to protect the privacy of its users. That same month, the project funders threatened to cease their funding if Facebook could not deliver data by September 30, 2019.¹¹⁴ The Social Science Research Council issued a statement outlining the steps they would take, which included pausing the review process and paying out full grants to current researchers regardless of data availability.¹¹⁵ Since August, Facebook has released some additional data, although not enough to enable grantees to complete their research. SSRC’s statement included the intention to wind down the project by the end of 2019. A December 2019 statement from Social Science One’s co-chairs and European Advisory Committee said that the data Facebook had

110. “Social Science One: Building Industry-Academic Partnerships,” Social Science One, <https://socialscience.one>, accessed September 13, 2019.

111. “Fundors,” Social Science One, <https://socialscience.one/funders>, accessed September 13, 2019. Full details on the organizational setup of Social Science One are available at Gary King and Nathaniel Persily, “A New Model for Industry–Academic Partnerships,” *Political Science and Politics*, August 2019, <https://gking.harvard.edu/files/gking/files/partnerships.pdf>.

112. Elliot Schrage, “Facebook Launches New Initiative to Help Scholars Assess Social Media’s Impact on Elections,” Facebook, April 9, 2018, <https://about.fb.com/news/2018/04/new-elections-initiative/>.

113. “Social Media and Democracy Research Grants,” Social Science Research Council, <https://www.ssrc.org/fellowships/view/social-media-and-democracy-research-grants/grantees>, accessed September 13, 2019.

114. Funders Supporting Independent Scholarly Access to Facebook Data, “Letter to The Social Science Research Council,” Social Science Research Council, August 27, 2019, https://ssrc-static.s3.amazonaws.com/sdi/resources/SMDRG_funder_letter_august_2019.pdf.

115. “Statement from Social Science Research Council President Alondra Nelson on the Social Media and Democracy Research Grants Program,” Social Science Research Council, August 27, 2019, <https://www.ssrc.org/programs/view/social-data-initiative/sdi-statement-august-2019/>.

released to that point was of “extremely limited scientific value,” and that “there is good reason to doubt whether other useful data will be forthcoming.”¹¹⁶ Social Science One’s Facebook partnership highlights the hurdles researchers face when working with private platforms, and also points to the difficulty that firms have in making data available to researchers when that data are drawn from users in countries with widely varying legal regulations around privacy protection.

Outside of Social Science One, some individual researchers have formed their own platform partnerships, though these arrangements are rare. For example, Politico reported back in 2018 that Facebook provided some de-identified user data to a team led by economist Raj Chetty for research on income inequality in the United States.¹¹⁷ Many different individuals and organizations have laid out their own projects and recommendations for solving data access issues for social science researchers. In 2017, the Commission on Evidence-Based Policymaking recommended establishing a National Secure Data Service that would link federal database systems and streamline access to government data.¹¹⁸ Though other recommendations from the Commission were enacted with the passage of the Foundations for Evidence-Based Policymaking Act of 2018 (H.R. 4174),¹¹⁹ the final bill did not include the creation of a National Secure Data Service.

Robert M. Groves, former Census Bureau director and current Georgetown University provost, and Adam Neufeld, vice president of Innovation and Strategy at the Anti-Defamation League, outlined a model similar to Social Science One’s that would make private-sector data available through an intermediary institution that could ensure privacy.¹²⁰ The Alfred P.

116. “Public Statement from the Co-Chairs and European Advisory Committee of Social Science One,” Social Science One blog, December 11, 2019, <https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one>.

117. Nancy Scola, “Facebook’s Next Project: American Inequality,” *Politico*, February 19, 2018, <https://www.politico.com/story/2018/02/19/facebook-inequality-stanford-417093>.

118. Nick Hart and Kody Carmody, “Barriers to Using Government Data: Extended Analysis of the U.S. Commission on Evidence-Based Policymaking’s Survey of Federal Agencies and Offices,” Commission on Evidence-Based Policymaking, October 2018, <https://bipartisanpolicy.org/wp-content/uploads/2018/10/Barriers-to-Using-Government-Data.pdf>.

119. “The President Signs H.R. 4174, ‘Foundations for Evidence-Based Policymaking Act of 2018,’” Social Security Administration, February 15, 2019, https://www.ssa.gov/legislation/legis_bulletin_021519.html.

120. Robert M. Groves and Adam Neufeld, “Accelerating the Sharing of Data Across Sectors to Advance the Common Good,” Beeck Center for Social Impact and Innovation at Georgetown University, 2017, <https://beeckcenter.georgetown.edu/wp-content/uploads/2019/07/Accelerating-Sharing-of-Data.pdf>.

Sloan Foundation's Administrative Data Research Facilities Network also offers its own data sharing model.¹²¹

Regulation and Legal Solutions

Experts had mixed opinions on what role, if any, regulation should play in solving challenges in the field. While some said that without regulation, digital inequalities will continue to perpetuate, others expressed concern about the unintended consequences regulations might have and skepticism as to whether laws around technology, many of which are crafted without input from the research community, would be the most effective means of solving problems researchers face. Several spoke of a greater need for researchers to be involved in the lawmaking process and to ensure that their voices are heard when writing legislation around data use.

“Legitimate public concern about privacy could easily be pitted against the legitimate interests of researchers,” said Christian Sandvig. “We need to ensure that these two goals are not seen as a trade-off for opposites.”

Many experts said that regulation could potentially play a vital role in providing data access and transparency that could transform research in this field. Some specifically cited changing provisions within the Computer Fraud and Abuse Act (CFAA),¹²² an anti-hacking law that broadly prohibits users from accessing computers without permission. The highly controversial law as it is currently written has implications for researchers¹²³ and makes terms of service violations criminal acts that can carry fines and prison sentences.¹²⁴

Several individuals and organizations have challenged the law in court, but a few recent cases have given experts some hope that change could potentially be on the horizon. Earlier this year, the Ninth Circuit Court of Appeals ruled in the *hiQ* versus LinkedIn case that scraping public data

121. Administrative Data Research Facilities Network, <https://www.adrf.upenn.edu>, accessed September 11, 2019.

122. “18 U.S. Code § 1030. Fraud and Related Activity in Connection with Computers,” Cornell Law School Legal Information Institute, <https://www.law.cornell.edu/uscode/text/18/1030>, accessed October 4, 2019.

123. Kim Zetter, *Wired*, June 29, 2016.

124. “Prosecuting Computer Crimes,” Office of Legal Education Executive Office for United States Attorneys, January 14, 2015, <https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf>.

“likely” does not violate the Computer Fraud and Abuse Act.¹²⁵ The Electronic Frontier Foundation hailed the ruling as “a major win for research and innovation” and “an important step to putting limits on using the CFAA to intimidate researchers with the legalese of cease and desist letters.”¹²⁶

Christian Sandvig along with a team of researchers, journalists, and the American Civil Liberties Union is also challenging the CFAA.¹²⁷ Filed in the U.S. District Court for the District of Columbia in June 2016, the case argues that the CFAA and overreaching terms of service agreements inhibit robust algorithm auditing and prevent users from uncovering discrimination.¹²⁸

Other experts called for better regulations that promote technology transparency. Users should know, for example, when a post is a paid advertisement, whether it was created by a human or bot, how users were targeted to see that post, and who is behind ad campaigns. Facebook has taken some steps toward transparency for ads about political and social issues,¹²⁹ though critics who were not interviewed for this report have pointed to the company allowing ads that contain false accusations as a move away from transparency.¹³⁰ Some experts said that transparency should apply to all platform activity, not just political speech. This does not mean eliminating anonymous speech, but instead making it clear that the origins of specific messages are unverifiable.

Several experts noted that regulations within the United States will most likely have limited efficacy as platforms continue to grow their base of international users. David Lazer pointed to a report for a further discussion

125. Judge Berzon, “hiQ Labs, Inc. vs. LinkedIn Corporation Opinion,” United States Court of Appeals for the Ninth Circuit, September 9, 2019, <http://cdn.ca9.uscourts.gov/datastore/opinions/2019/09/09/17-16783.pdf>.

126. Camille Fischer and Andrew Crocker, “Victory! Ruling in hiQ v. LinkedIn Protects Scraping of Public Data,” Electronic Frontier Foundation, September 10, 2019, <https://www.eff.org/deeplinks/2019/09/victory-ruling-hiq-v-linkedin-protects-scraping-public-data>.

127. “Sandvig v. Barr—Challenge to the CFAA Prohibition on Uncovering Racial Discrimination Online,” American Civil Liberties Union, May 22, 2019, <https://www.aclu.org/cases/sandvig-v-barr-challenge-cfaa-prohibition-uncovering-racial-discrimination-online?redirect=cases/sandvig-v-sessions-challenge-cfaa-prohibition-uncovering-racial-discrimination-online>.

128. Documentation from the case is available online at *ibid*.

129. Katie Harbath, “Updates to Ads About Social Issues, Elections or Politics in the US,” Facebook, August 28, 2019, <https://about.fb.com/news/2019/08/updates-to-ads-about-social-issues-elections-or-politics-in-the-us/>.

130. Cecilia Kang, “Facebook’s Hands-Off Approach to Political Speech Gets Impeachment Test,” *New York Times*, October 8, 2019, <https://www.nytimes.com/2019/10/08/technology/facebook-trump-biden-ad.html>.

of concerns independent academics who were not interviewed for this report have with digital platforms and the challenges of creating policies to solve those issues.¹³¹

131. “Stigler Committee on Digital Platforms Final Report,” Stigler Center for the Study of the Economy and the State, 2019, <https://www.publicknowledge.org/wp-content/uploads/2019/09/Stigler-Committee-on-Digital-Platforms-Final-Report.pdf>.

Acknowledgments

Special thanks to all of the experts interviewed for this report:

Jessica Baldwin-Philippi, associate professor in the Communication and Media Studies Department at Fordham University

Henry E. Brady, dean of the Goldman School of Public Policy and Class of 1941 Monroe Deutsch Professor of Political Science and Public Policy at the University of California, Berkeley

Joan Donovan, director of the Technology and Social Change Research Project at Harvard Kennedy School's Shorenstein Center on Media Politics and Public Policy

Deen Freelon, associate professor in the Hussman School of Journalism and Media at the University of North Carolina at Chapel Hill

Sam Gill, vice president of Communities and Impact and senior advisor to the president at the John S. and James L. Knight Foundation

Gary King, Albert J. Weatherhead III University Professor at Harvard University, director of the Institute for Quantitative Social Science, and director of Social Science One

Kate Klonick, assistant professor of law at St. John's University and affiliate fellow at the Information Society Project at Yale Law School and New America

David Lazer, professor of political science and computer and information science at Northeastern University and co-director of the NULab for Texts, Maps, and Networks

Jonathan Nagler, co-director of the New York University Social Media and Political Participation (SMAPP) Lab and director of the NYU Politics Data Center

Alondra Nelson, president of the Social Science Research Council and Harold F. Linder Chair in the School of Social Science at the Institute for Advanced Study

Christian Sandvig, professor of digital media at the University of Michigan and director of ESC: the Center for Ethics, Society, and Computing

Jen Schradie, assistant professor at the Observatoire Sociologique du Changement (OSC) at Sciences Po in Paris

Aaron Smith, director of Data Labs at the Pew Research Center

Francesca Tripodi, assistant professor of sociology at James Madison University and affiliated researcher with the Data & Society Research Institute

Ethan Zuckerman, associate professor of the Practice in Media Arts and Sciences at the Massachusetts Institute of Technology (MIT) and director of the MIT Center for Civic Media

About the Author

Christina Couch is a science and technology journalist and Assistant Director of Professional Development for the MIT Graduate Program in Science Writing. Her work can be found in *The New York Times*, *NOVA Next*, *MIT Technology Review*, *Fast Company*, *Science Friday*, and *The Verge*.

American Academy of Arts & Sciences

Board of Directors

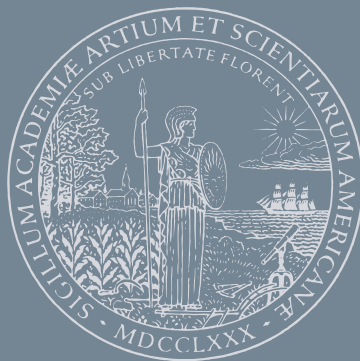
Nancy C. Andrews, *Chair*
David W. Oxtoby, *President*
Alan M. Dachs, *Vice Chair*
Diane P. Wood, *Vice Chair*
Carl H. Pforzheimer III, *Treasurer*
Geraldine L. Richmond, *Secretary*
K. Anthony Appiah
Louise H. Bryson
John Mark Hansen
Nannerl O. Keohane
John Lithgow
Cherry A. Murray
Venkatesh Narayanamurti
David M. Rubenstein
Larry J. Shapiro
Shirley M. Tilghman
Natasha Trethewey
Pauline Yu

Selected Publications of the American Academy

David Karpf, *The Internet and Engaged Citizenship* (2019)

Caroline B. Brettell, *The Political and Civic Engagement of Immigrants*
(2020)

To order any of these publications please contact the Academy's
Publications Office. 617-576-5085; publications@amacad.org



AMERICAN ACADEMY OF ARTS & SCIENCES

 @americanacad

www.amacad.org