Research Coordination Network: Democracy in the Networked Era

The Digital Information Environment & Global Elections

Authors

Cody Buntain, Josephine Lukito, Jacob N. Shapiro, Lisa Singh, Dhanaraj Thakur, Cristina de la Puerta, Zeve Sanderson, Joshua A. Tucker[1]

Acknowledgments

We would like to express our gratitude to NSF's *Secure and Trustworthy Cyberspace* (SaT-C) program, which has funded this Research Coordination Network and its programming (NSF Award #2331641). We would also like to thank the attendees of the Digital Information Environment & Global Elections in 2024 conference, whose insights contributed to this report.

Introduction

In May 2024, the NYU Center for Social Media & Politics hosted the inaugural event of a new Research Coordination Network (RCN): Democracy in the Networked Era. With the support of the NSF's *Secure and Trustworthy Cyberspace* (SaT-C) program, this network was formed with the goal of fostering an interdisciplinary exchange to advance research on the online information ecosystem, including important goals for the field moving forward, the types of infrastructure needed to support this work, and methods for translating insights from the research community to policymakers and vice versa.

Our first research meeting focused on *The Digital Information Environment & Global Elections in 2024* and brought together scholars from computational social science, data science,

[1] The first draft of Section I was written by Josephine Lukito, Section II by Cody Buntain and Jacob N. Shapiro, Section III by Lisa Singh, and Section IV by Dhanaraj Thakur. The first draft of the Introduction, Conclusion, and Executive Summary were written by Cristina de la Puerta. All authors contributed to the editing of this report. Joshua A. Tucker is the Principal Investigator on NSF Grant 2331641 which funded the conference and white paper.

human-computer interaction, and computer science, among other disciplines, to share insights across this field of study and address pressing questions amidst a rapidly changing digital media environment.

This report summarizes two days of research presentations and workshops addressing four key areas that participants identified as being of the highest priority to the field of research on digital media and democracy in an unprecedented year for elections globally.

Executive Summary

During *The Digital Information Environment & Global Elections in 2024* research meeting, participants identified four main areas of concern that needed to be addressed in the field of digital media and democracy research: research infrastructure; baselines and tooling in experimental research; sampling and observational data; and generative AI. This section summarizes the key questions and opportunities for collaboration that attendees identified in order to better understand these issues, as well as takeaways and recommendations from each section; the rest of this report addresses each of these topics in much greater detail.

Section I: Research Infrastructure

RCN participants recognized a need for improving data infrastructure and identified a number of recommendations that would bolster data collection methods, understanding of data ethics, and data accessibility in the field. Research infrastructure is essential for consistent, robust, and independent scholarship on the digital information environment. At present, however, there are few incentives to build and support research infrastructure, particularly in the area of data access and collections. While researchers have long used social media data to study content virality, the spread of information, and the relationship between online and offline behaviors, the infrastructure needed to support this research has become increasingly fragmented. Especially as existing APIs like X/Twitter’s and Meta’s Crowdtangle have disappeared or become less accessible, researchers need more support than ever before to access essential data.

Within this context, many scholars have turned to alternative methods of data collection, such as data donations and web scraping, each of which comes with their own technical and/or legal challenges. Data donations rely on users voluntarily providing their data, a method that offers platform users an explicit option as to whether or not to consent to having their data included in the study, but is also accompanied by challenges, such as data security concerns and selection bias based on who is willing to donate data (and especially the extent to which this may be related to digital literacy or political ideology). Web scraping, while increasingly popular, operates in a legal gray area, making it an often unstable solution for scholars. Both options can also be disrupted by changes made by the platforms.

With new challenges emerging—and past challenges intensifying—for collecting large-scale digital data, we identified opportunities for collaborative investments in shared infrastructure. Some of the key takeaways include:

- **Multi-national, Multi-institutional Collaborations**: Developing "many-lab" infrastructures with cross-institutional IRBs and shared grants would enable global cooperation and increase researchers' access to support.

- **Shared Datasets**: Open-access datasets should be prioritized for comparing different platforms or time periods, benefiting under-resourced researchers and enabling consistent comparisons.

- **Archival, Ethical, and Security Standards**: Established standards for data encryption, anonymization, and sharing that aligns with ethical guidelines and can adjust to different data types can help ensure safe and secure data storage, and thus facilitate greater sharing of data across research groups.

- **Data Storage & Access**: Implementing a centralized system to vet researcher access, connect research teams, and provide storage space to encourage data reuse and infrastructure sustainability.

- **Policies for Responsible Access**: Advocating for policies that support independent research while balancing user privacy, such as those included in the EU's Digital Services Act, to be adapted in the U.S. where many major tech companies are based.

Section II: Experimental Research

Workshop participants also explored the role of baselines and standardized tooling in improving researchers' ability to use experimental analyses to better understand behavior in the digital information environment. Here, baselines refer to measures of the normal state of users and/or content on a platform. Experiments allow researchers to observe how individuals respond to changes in specific elements of their digital media experiences, such as deactivating platform features or introducing new content. This section of the report explores a variety of different experimental methods, including survey, lab, and field experiments. It also examines new tools for experimental research, such as browser extensions and bespoke research-oriented platforms, designed to enable platform modifications.

This section also highlights the importance of designing experiments that allow for the study of particular sub-populations. Participants noted the possibility that certain experimental interventions might not have an effect on an average person, but might still prove consequential for important sub-portions of the population, such as people with exposure to particular parts of the information environment. More technically, interventions might generate null effects when average treatment effects are measured across full populations even when there may be important effects on smaller subpopulations. For this reason, we argue for the importance of research focused on particular population sub-groups, though engaging these populations presents its own set of challenges, including distrust in scientific research by some communities.

Experiments are also critical for understanding the impact of interventions online. In particular, attendees noted that further research would be helpful for understanding how to design platform features that encourage cross-cutting conversations. Experimental studies can also further our understanding of content moderation policies and practices, and how they affect users' motivations for sharing information or, potentially, lead them to turn away from certain conversations and platforms.

Key takeaways from this section include:

- **Expanding Baseline Data**: Developing comprehensive baselines (measurements of the normal state of users and/or content on a platform) for social media activity can allow researchers to better target sub-populations and create more externally valid studies.
- **Developing Standardized Tools**: Investing in browser extensions, bespoke platforms, and generative agents can facilitate more – and more comparative – experimental research by decreasing barriers to entry.

Section III: Sampling & Observational Data

In addition to experimental research methods, workshop participants identified a need in the field for sampling baselines that would be beneficial for researchers to better understand the relationship between digital media and democracy, especially as it pertains to conducting observational studies.

Observational studies are crucial to our understanding of public opinion, behavior on digital platforms, and the shifting dynamics of the information environment. Central to these studies is the need for reliable baselines or benchmarks, which provide researchers with reference points for meaningful analysis. The challenge, however, is that efforts to establish benchmarks—particularly for election-related studies—have been fragmented and isolated.

To address these challenges, RCN participants advocated for collaborative and globally-organized benchmark efforts to enhance the value of observational research, especially in understanding how different platforms function during election campaigns and their aftermath. They emphasize that cross-platform and multi-modal comparisons are essential for creating a more comprehensive view of social media usage during elections, especially as our digital media environment continues to fragment and expand beyond traditional media.

By creating shared, robust baselines, researchers can improve the quality of their studies and ensure a more accurate representation of digital and election dynamics across different platforms. This includes random sets of user-based and content-based samples. These baseline samples can be used to understand the prevalence of different types of content, enabling researchers to specifically study how election-related conversations evolve online.

Key takeaways from this section include:

- **Establishment of Baseline Platform Content**: Descriptive statistics like platform usage time, demographic details, languages, and user activity are needed to form meaningful benchmarks.
- **Election-Specific Data**: Unique baselines for elections, such as threat assessments and political elite behaviors, are important for monitoring election integrity.
- **Collaboration and Global Standards**: Harmonizing existing efforts through cross-national surveys and collaborative institutes to make data and methodologies widely available.
- **Measurement Comparison**: Establishing infrastructure for comparing computational methods used for different measurements, e.g. state of the art content detection methods.

Section IV: Generative AI

The rapid rise and accessibility of generative AI models has surfaced questions across disciplines about the impact these tools could have on elections, both positive and negative, and how they could reshape what we know from previous research. Participants noted a few main areas of interest as it pertains to generative AI's potential impact on the information environment and elections: voter access to information, foreign influence operations, and political campaign use, as well as a number of additional, more nuanced applications.

In terms of voters' access to information, AI tools have the potential to lower the cost of creating content, and therefore increase the scale of content creation. A key concern identified was how foreign influence campaigns could use the new resources at their disposal to quickly create realistic content in a wide variety of media types (text, image, audio, video, and multimedia). Further research is necessary to understand the persuasiveness of generated or augmented content in different forms. Further research is also needed to understand how people interpret AI produced content – e.g., do they believe it to be produced by AI, humans, or some combination of both? – and the impact of *labeling* AI generated content on these beliefs. Relatedly, it is important to understand whether users' exposure to and/or awareness of such labels changes how users interpret *unlabeled* content.

Chatbots and AI-informed search engine results also create a new avenue for voters to seek information about politics. Greater transparency around users' political inquiries on these platforms would be beneficial for researchers to understand how voters are using them in practice, as well as the quality of responses that these resources produce. Generative AI also presents opportunities for less resourced down-ballot campaigns to make content that might have previously required significant financial or labor investments. For example, candidates now have the ability to create materials in foreign languages to reach new populations of voters.

Finally, participants noted that conversations around AI's impact require tremendous nuance, especially as understanding the use of AI in politics involves combining rapidly evolving technology with understanding both how humans utilize that technology for political purposes and how humans respond to interactions with AI in a political context. Not only do these tools create a new context in which we now need to understand previous research, but they also present a wide variety of uses and implications in a myriad scenarios, as well as second-order effects which this section discusses in more detail. Female politicians and journalists have also often been the subject of targeted deepfake attacks, demonstrating the importance of attention to subpopulations in understanding AI's impact.

Key takeaways from this section include:

- **Improving Transparency**: More transparency is needed to understand how AI models are used, particularly to address biases and disparate impacts on different demographic groups. Initiatives such as data donations from users can help address this gap.
- **New Frameworks for Previous Findings**: Previous findings on the spread and persuasiveness of information can also help researchers prioritize and inform approaches for understanding its impact on elections.
- **Further Research on Interventions**: In order to create informed policies on AI and its second-order effects, like declines in trust, further research is needed on potential interventions like inoculation and labeling of AI content in different forms.

Section I: Research Infrastructure[2]

Introduction

Across multiple disciplines, many researchers have relied on social media data as a proxy for other social phenomena or as an object of study itself. This includes studying platforms such as NextDoor (Brown et al., 2024), X/Twitter (Jungherr 2015), and TikTok (Primig et al., 2023) to understand content virality, the spread of mis/disinformation and polarizing content (Rathje et al., 2023), and the relationship between online and offline behaviors (Lukito et al., 2023).

To conduct this work, researchers need access to technical, administrative, legal, and ethical infrastructure to produce and sustain high-quality scholarship. Thus far, infrastructure in this area has been built piecemeal, with individual researchers and research teams primarily building bespoke infrastructure for specific projects or needs. There is an increasing demand for more widely accessible traditional infrastructure such as programming packages, open-source tools, and means for archiving data and code. However, there is also growing recognition about the need for support in navigating other challenges related to researching the digital information environment, including ethical and legal support. Shared infrastructure not only elevates the robustness of research, but it also democratizes access to infrastructure for early career scholars and researchers who are not as well-resourced to build their own infrastructure. In doing so, investments in research infrastructure can not only benefit the research community, but also can subsequently lead to research that can inform policymaking and civic knowledge.

Of the plethora of different research infrastructure necessary to conduct platform research, data access infrastructure has become a hot button topic. In this section, we will explore current methods for collecting social media data, limits to contemporary collection strategies, and proposed solutions for shared infrastructure that may benefit the research community.

Current Data Collection Strategies

There are several approaches to collecting data from the digital information environment. While research from the past decade has benefitted from more widely accessible, and free, research Application Programming Interfaces (APIs), many of these data access points are no longer available. For instance, the X/Twitter API, which had long been free – or relatively inexpensive – for academics to use, is now prohibitively expensive for researchers. Other tools, such as Meta's Crowdtangle platform, which was used to download data from public accounts on Facebook and Instagram, shut down access on August 14, 2024. And archives developed by researchers, such as Pushshift (Baumgartner et al., 2020), have disappeared under pressure from platforms. This makes it increasingly challenging to study changes in the digital information environment, or effects on individuals and groups since **these traditional sources of data are becoming (or have become) inaccessible**.

[2] This first draft of this section was written by Josephine Lukito.

In the wake of the API closures, there has also been a growing call for policy and regulations that provide researchers with API access to conduct independent research on social media platforms. While there are some optimistic developments (such as the development of Article 40 of the General Data Protection Regulation in the European Union [EDMO 2023]), the practical implementations of these policies are still in development.

Given these closures, researchers have also turned to alternative data collection strategies that request data directly from citizens and social media users. For example, one increasingly popular method is the use of donated data. Data donation research uses digital trace data that is voluntarily provided by consenting research participants. This can be done in one of two ways: by **requesting existing data** from participants, or by **using browser extensions** and other software to collect users' data in real time (Araujo 2022).

There are several key advantages of this method, including that consent is transparently and directly provided by the user, and researchers are able to study both **content consumption** and **content production** (whereas API-provided data has primarily focused on content production). Data donation projects can also be used to conduct experiments and audits. For example, the project Intervenr utilizes a reusable infrastructure that includes a browser extension, a web application, and a back-end that collects advertisements targeted to research participants (Lam 2023). Embedded in this infrastructure is a data redaction interface that also allows research participants to exclude or anonymize data from the study.

However, there are also challenges to data donation work. For one, people are not necessarily incentivized to share their data or may be unfamiliar with how to provide their historical user data. For many platforms, this process can also be opaque and time-consuming for both users and researchers. Additionally, because of the potential sensitivity of donated data, particularly from private citizens, it is necessary to develop more consistent standards and practices for the storage and security of this data, which may make it more challenging to pool or share data amongst researchers.[3] Further, inferences drawn from studies based on data donation need to wrestle with the question of the representativeness of samples of users willing to donate data to other populations of interest.

Another data collection method that researchers are once again utilizing is web scraping or extraction techniques, which involve the use of programming tools or software to extract data or content from publicly accessible websites and apps. Often, when researchers do not have access to an API, they will turn to **data scraping** as an alternative for gathering content.

But these methods exist in a legal gray area. Prior cases in the United States, such as HiQ v. LinkedIn (2021), suggest that data scraping is not considered a violation of the Computer Fraud and Abuse Act (CFAA). However, these court rulings have not stopped platforms from (unsuccessfully) attempting to sue other researchers who are scraping their platforms.[4]

---

[3] It is worth noting that some research centers, such as the National Internet Observatory, are working to build such infrastructure: https://nationalinternetobservatory.org/
[4] See for example lawsuits filed in 2023 against Bright Data and Center for Countering Digital Hate (CCDH). Both cases were dismissed.

Cumulatively, these current data collection strategies highlight the ongoing precarity of research on the digital information environment. What researchers did to collect data a decade ago cannot be replicated today as data access regimes are constantly changing – and we should expect this to continue into the future. This is not a feasible approach for continuing research: to produce sustainable research with robust findings, researchers must have **consistent access to reliable data infrastructure**.

Motivating Incentives for Infrastructure Building

As noted above, there are few incentives to support data access infrastructure. Because the study of the digital information environment is interdisciplinary, researchers may be motivated by different goals. For example, whereas infrastructure itself may be considered a scholarly contribution to computer science, social science researchers may outsource these efforts for specific projects because they are not as incentivized to build sustainable data infrastructure. Furthermore, these norms also incentivize the development of new infrastructure, rather than the **maintenance of existing infrastructure**, because new work brings publications whereas maintenance is not as equally rewarded.

Yet another challenge is that in some fields academic competition also disincentivizes the sharing of infrastructure, as sharing data may mean being poached on a research project or idea. Because of this, researchers are more likely to all build their own data collection methods individually, rather than **collaborating on shared infrastructure** that can benefit the field as a whole.

To motivate incentives for infrastructure building, two developments are essential. First, the field must specify what infrastructural needs would most benefit the research community. This white paper attempts a first step in this regard in the following section. Second, researchers, funders, policymakers, and citizens must work together to build and support infrastructure for studying the digital information environment. This is not an endeavor that can be solely taken by researchers—**other stakeholders must be involved** to ensure that this research is **transparently conducted, well-supported, and benefits the public good**.

Conclusion & Needs of the Field

In this report, we identify four (4) key needs that would benefit the development of research infrastructure for the study of the digital information environment generally, in addition to the digital information environment in the context of elections. Importantly, we acknowledge that there may be other needs, particularly as the field develops.[5] As data access remains a precarious endeavor, we focus especially on suggestions that would benefit joint data collection strategies, multi-institutional collaboration, ethical data sharing, and data security norms.

Multi-National, Multi-Institutional Collaborations. One pivotal need of the field is the development of “many-lab” infrastructure, or research infrastructure that can support

[5] We are also far from the first group of researchers to consider these types of questions; see for example https://www.smrconverge.org/home/methodology/.

multi-institutional collaborations. Such efforts have been made in other fields, including in the psychological sciences (Psychological Science Accelerator, 2017) and in policy research (Evidence in Governance and Politics), to much success. Internet researchers are increasingly collaborating with those in other institutions, or with non-academic researchers, even without incentives or support; we can expect such endeavors to increase, rather than decrease, in the future.

For the study of digital media platforms, multi-institutional infrastructure includes the development of **cross-institutional IRBs, multi-institutional grant support,** and **financial incentives for multi-team collaborations**. Multi-institutional collaborations may also be able to collectively bargain for infrastructural support and for funds to support ongoing, rather than new, projects.

Shared Datasets. Another need of the field are open-access and shared datasets, particularly baseline datasets that can be used to develop and compare new and old tools or platforms. For example, comparisons of content across different platforms (e.g., YouTube and TikTok) or across different time spans are predicated on having comparable datasets. Fields with existing archives of datasets often benefit from being able to compare findings across different papers and projects that use the same dataset, and can also make data more accessible to under-resourced researchers. Here, again, we hardly need to invent something new, as fields such as electrical engineering (IEEE 2019) and public opinion (Roper Center for Public Opinion Research) have long relied on shared datasets to conduct research. In fact, the Inter-university Consortium for Political and Social Research (ICPSR), which maintains one of the longest standing data archives for social science, continues to lead the field in this area with the development of their Social Media Archive (SOMAR). Other more targeted efforts in this space include: Junkpedia and the Massive Data Institute (MDI) text analytic portals.[6]

Archival, Ethical, and Security Data Collection Standards. The third data infrastructure needs we propose is, broadly, a set of standards or norms around data storage/archiving and data sharing. Even without being incentivized to do so, researchers continue to share data for the benefit of the research community and the public[7]— but a lack of norms regarding the "right" way to share data hinders such efforts.

We therefore recommend a shared set of field-wide standards regarding encryption, data anonymization procedures, and data sharing practices, including a standard data use agreement or data release policy. These routines should adhere to ethical expectations as proposed by professional organizations, including the Association of Internet Researchers (AOIR).[8] Importantly, different standards may need to be adjusted for different kinds of data; for example, content from public figures that is collected via an API or data scraping tool has different privacy concerns

---

[6] See for example Lisa Singh, Colton Padden, Pam Davis-Kean, Rabin David, Virinche Marwadi, Yiqing Ren, and Rebecca Vanarsdall (2021). Text analytic research portals: supporting large-scale social science research (Poster). IEEE International Conference on Big Data, online.

[7] For example, Patrick Warren and Darren Linvill collaborated with FiveThirtyEight to release a dataset of tweets from Russian troll accounts in 2018, making it possible for disinformation researchers to study this foreign interference campaign without relying on platform-provided data. For more, see https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/

[8] For more, see the Ethical Guidelines 3.0 of AOIR: https://aoir.org/reports/ethics3.pdf

compared to donated data provided by private individuals about their social media consumption habits.

Policy for Responsible Data Access Infrastructure. Last, but certainly not least, there is an essential need for policy that supports independent research on the digital information environment. Such policy should balance data access for researchers and privacy expectations of users. For example, in the European Union, data provided to researchers through Article 40 of the Digital Services Act must still adhere to the General Data Protection Regulation. Similar policy efforts can and should be made within the United States.

Section II: Experimental Research[9]

Introduction & Defining "Experiments"

We define experiments as studies in which researchers randomize some treatment and compare outcomes between treated and untreated groups, or across multiple treatment arms. Broadly speaking there are three types of experiments that feature prominently in research on social media (Mosleh, Pennycook, & Rand 2022):

1. Field experiments, in which some aspect of subjects' social media experience is randomized. An example can be found in Coppock et al. (2015), which randomly exposed Twitter users to either a public tweet or a direct message from a non-profit requesting signatures on a political petition.

2. Lab experiments, in which researchers invite subjects to participate in a fully artificial environment and vary different aspects of their experiences, such as exposure to election (Allcott & Gentzkow, 2017) or COVID-19 related information (Pennycook et al., 2020).

3. Pure survey experiments where the treatment is delivered in a survey and survey questions are used as the key outcome measure (Moon et al, 2022).

Because many aspects of the social media environment are hard to replicate in artificial settings, researchers have become increasingly interested in field experiments over time. These have included: deactivation experiments in which subjects turn off a given platform (Allcott et al., 2024); changes to platform features, which are done all the time by major social media platforms A/B testing different features (Xu et al. 2017), and also sometimes by researchers through browser extensions (Robertson et al., 2018) or in collaboration with platforms (Nyhan et al., 2023); changes to how algorithms deliver content to users' feeds (Guess et al., 2023a); and injections of new content, such as experiments on the impact of content labels (Aslett et al., 2022).

Scope of Experiments

Experimental research with individual-level treatments provide unique value in addressing certain kinds of questions for which understanding causal impacts are essential.

When researchers want to understand the impact of individual-level interventions on behavior, field experiments are a very powerful tool. Workshop participants noted, however, that their value is lower for behaviors where network effects are an important incentive, because for such behaviors the change in the value of taking one action vs. another from an individual-level treatment is likely to be small.

[9] This first draft of this section was written by Cody Buntain & Jacob N. Shapiro.

Workshop participants also noted the particular value of experiments for studying **individual-level incentives**, i.e. what makes an individual choose a given action or not. In the context of current concerns around how the information environment is used for political conversations, workshop participants indicated an interest in questions such as:

- What kinds of platform features or feedback would encourage **cross-cutting conversations**?
- What kinds of features would encourage creation of **less extreme content** (either in terms of viewpoint or simply the kinds of argument made)?
- What motivates people to share new kinds of information, and which types of information **mobilize people to engage** on a topic for the first time?
- Which kinds of experiences lead people to **turn away from online engagement** in certain kinds of conversations, or from online spaces overall?

Workshop participants also identified a **tension in experimental designs that focus on average treatment effects versus effects in the tails** of the social media population. In the political context, experimental studies on various aspects of the social media experience, as well as theoretically motivated interventions to impact individual attitudes and behaviors, often return null findings for average treatment effects. While many experimental treatments may have null effects on the population writ large, these same treatments **may be impactful on particular sub-populations** of interest. Social science experimental research has a strong foundation of work on this issue around "Heterogeneous Treatment Effects" (HTEs), but we have reason to suspect that missing a particular set of HTEs may be particularly consequential in the study of online behaviors. Prior work on ad-targeting, for example, shows that relatively small audiences of conspiracy-theory consumers on YouTube receive a largely unique set of advertisements in the platform (Ballard et al., 2022). Starbird (2017) reveals a prominent role of alternative media and advertising in propagating conspiracies around mass shootings, and leaks from Meta illustrate how Instagram content about eating disorders is quickly shown to at-risk groups (Wells et al., 2021). As platforms already engage with these small—and potentially vulnerable—audiences, the academic community must similarly engage with these small populations.

This engagement with small, potentially at-risk populations is fraught in U.S. and Western contexts, where certain audiences may be particularly distrusting of scientific inquiry and perceived surveillance, e.g., the authors of Zannettou et al. (2017) were routinely harassed by members of 4chan. Issues arising from engagement with these sub-populations become more complex as experimental studies expand to focus on the Global South and low- and middle-income countries, where language- and cultural-specific expertise is necessary. Workshop participants identified recent efforts to engage with these global audiences: e.g., Carey et al. (2020) focus on Brazil; Javed et al. (2020) studied COVID-19 messaging in Pakistan; and a large body of work studying online information spaces in India has emerged in recent years. Workshop participants highlighted a critical need for **more work in the Global South**, as social media platforms and the internet are increasingly blended in these spaces, with unique challenges around national zero rating (wherein mobile-network operators privilege certain mobile applications by excluding data used by these applications from subscribers' data allowance, thereby incentivizing concentrated use among these zero-rated applications), in-platform authoritarian censorship, and governments' willingness to disable access to social media during moments of unrest (e.g., as in Sri Lanka [Freedom House, 2022], Iran [Khalaji 2022], and more recently in Bangladesh [Diya 2024]).

Finally, workshop participants noted that scholars often start addressing a topic in **highly artificial settings** and build to realism. They discussed the evolution of behavioral economics which moved from completely artificial lab experiments to complex natural field experiments, as well as work on accuracy nudges to reduce misinformation sharing, which started in the lab (Nyhan & Reifler, 2010) and have now moved into the field (Pennycook et al., 2021).

Baselines for experimental work

Workshop participants identified a number of ways that baseline data can support experimental work. We use the term baseline data to refer to measures of the normal state of users and/or content on a platform. These measures can be generated with continuously collected samples of activity in online environments that allow one to identify **how people consume, engage with, or produce content absent any experimental intervention**. Baselines could include: measures of content creation, such as daily collections of videos and comments on a random sample of 1,000 YouTube channels drawn from the population of channels with more than 100,000 subscribers; measures of user interest, such as daily snapshots of Google and Bing search results for a set of 100 health-related search terms; or measures of content consumption, such as URL visits on mobile devices for a representative sample of 1,500 likely voters in a country.

The first, and most obvious use for baselines in experiments, is providing sample frames that scholars can draw upon in designing experiments. Experimental studies are almost always done on small samples. When these are representative of a well-defined population, one can extrapolate experimental findings to the larger population, subject to standard concerns about external validity (Bo & Galiani, 2021).

For experimental research in fields such as public health and medicine, this extrapolation can often be done on readily observable demographic characteristics because we believe most medicines work similarly for people of the same age, gender, and ethnic background living in similar places. No such expectation exists for treatments in the online space because **how individuals respond to content can be highly dependent on a wide range of social and psychological factors**.

It would thus be extremely useful for online experimentation for scholars to be able to draw **samples that are representative** of those who post certain kinds of content, respond in certain ways to events, or have similar social networks. It would also be extremely valuable for scholars to be able to benchmark convenience samples or those chosen on the basis of demographics (e.g., through companies that maintain panels of willing respondents such as CloudResearch or SurveyMonkey) against representative samples on the basis of online activity. Having baseline data on typical activity on different platforms would make it possible for researchers to say which online populations could reasonably be expected to respond to treatments in the same manner as their experimental subjects.

These baselines can also greatly facilitate targeted experiments described above, as the two-step process needed to identify potential participants in tail-based populations can be bootstrapped with these baselines. With such readily available baselines, researchers could identify elites in these collections whose audiences comprise these sub-populations of interest and then sample from actors that follow, amplify, or otherwise engage with these elites. This process has the potential to **significantly reduce costs to engage with and support** these important and potentially vulnerable groups.

Baselines would also be extremely valuable for developing treatments to use in natural field experiments, which aim to vary respondents' experiences in highly realistic settings. In economics, it has long been established that there can be a large difference between decision making in highly-artificial lab settings and the choices people make in real-world settings. In the context of research on social media, such studies may involve randomizing subjects into receiving different kinds of messages on real platforms (Munger et al., 2017) to see if survey experiment results on behaviors such as information sharing replicate, or having subjects interact with automata that respond using models trained on real-world data (Monstead et al., 2017)—a method that is particularly salient given advancements in large-language models (Törnberg et al., 2023). For such studies, having samples of activity that were drawn from well-defined populations, i.e. baseline data, would enable researchers to develop more realistic treatments.

Tools for running experiments

Beyond baseline data, workshop participants identified four types of tools that would enable a wide range of experimental research.

Browser extensions that allow for content to be inserted on common social media sites or in search as part of an experimental treatment would facilitate deployment of experimental manipulations directly within a user's browsing environment. By providing researchers with evidence on the effects of these manipulations in a real-world context, such tools would enhance the external validity of studies, as well as potentially by collecting aspects of user behavior pre- and post-treatment. Scholars have used browser extensions to study news consumption (Kleppe & Otte, 2017), the effectiveness of nudges towards politically balanced media (Munson et al., 2013), and incidental news exposure online (Möller et al., 2019). Workshop participants were not aware of any centrally maintained set of extensions that researchers can leverage to rapidly deploy in-browser treatments. Instead, research groups typically have graduate students or postdoctoral scholars develop one-off software for each study (sometimes building on open source tools). It would be more efficient for the field to maintain a few standard toolkits for running such experiments on common browsers, such as Google Chrome.[10]

[10] Similarly, the process of getting these extensions approved by Google for Google Chrome remains a one-off process between Google and whatever teams producing the extension. A centralized mechanism for facilitating the approval of extensions for academic research would also go a long way towards making the standup of such studies a more efficient process.

Scholars have advocated for a number of years for the development of **bespoke, research-oriented social media platforms** (Bail et al., 2023). These would allow researchers to vary core design features and measure outcomes using the kinds of fine-grained data on user behavior that is currently only available to researchers working inside major technology companies (e.g. dwell time on images). Such platforms would enable researchers to vary platform features such as interface design, algorithms for ordering content, or the availability of different options for interacting with other users. The key challenge in developing such platforms is getting a large enough user base that behavior will be reasonably close to what happens on large commercial platforms. Workshop participants noted that one or two such platforms might be sustainable with a combination of compensation and appeals to the social value of contributing to knowledge development. Others thought there was greater potential for using LLM-powered simulated personas to populate artificial social media environments (e.g., Törnberg et al. 2023). Either way, research will be needed to determine the external validity of such bespoke research platforms.

A third area for development highlighted in workshop conversations was **tooling for opt-in data collection**. With the erosion of API-based access on Facebook and Twitter, legal challenges to scraping in some jurisdictions, and the movement of most activity to mobile devices, it is increasingly important to facilitate data donations from users. Such donations would facilitate interesting new modes of experimentation such as the causal impact of real-world experiences (which can be varied through field experiments) on online behavior. While there are some tools for opt-in data collection (Melo et al., 2019), workshop participants noted these can be hard to use and are rarely maintained for long, meaning each research group that wants to use opt-in collection needs to spend significant resources on getting tools up and running.

Workshop participants also noted that much of these frameworks for data collection have focused on the information consumer even though information producers, such as political/media elites and content creators, definitionally have an outsized influence on these spaces, especially for the study of politics. Few **tools and resources are readily available to support these producers and content creators to share the analytics data they use to guide their creation process.** Growing our understanding of this cohort of creators could be massively improved through relatively small investments in tools for allowing these groups to donate the analytic data they use to make these decisions. Through the development of such tools, researchers could get finer-grained insights into cross-platform content creation strategies, thereby addressing open questions about the larger ecological environment when a single platform institutes policy changes.

Finally, workshop participants encouraged the development of **open-source generative agents trained on real-world social media data** (e.g., the baseline datasets mentioned above) that could support researchers in two ways. First, by providing the ability to rapidly pre-test treatments, generative agents would enable much faster research development cycles and save substantial costs from pre-testing with real people. Second, by enabling treatments in which respondents are randomized into having different kinds of interactions with a simulated entity, e.g. receiving polite criticism on a post vs. caustic criticism, generative agents could facilitate more naturalistic experimental studies, reducing external validity concerns.

Conclusion

There are many ways in which baseline data and standard tools could enable experimental research. Workshop participants noted significant complementarities between them. For example, baseline data can be used to improve the **external validity** of studies in multiple ways: by providing **training data for generative agents** that speed pre-testing and facilitate administering interactive **treatments at large scale**; by allowing scholars to recruit representative samples; and by enabling scholars to benchmark experimentally-induced variation against **behavioral differences** between populations in the real world.

Section III: Sampling & Observational Data[11]

Introduction

In order to characterize the world we live in and how it is changing, we need to continue to conduct observational studies. Observational studies give insight into public opinion, behavior of different subpopulations online, and the changing dynamics of our information environment across different platforms. However, to conduct observational studies well, researchers need access to a set of benchmarks or baseline statistics against which to compare new findings from their studies. To date, efforts to build social media benchmarks have been fairly independent, and especially so for benchmarks related to elections. In this section of the white paper, we identify the types of baselines needed for improving our understanding of elections and then argue that these should be joint, organized efforts that can be used by researchers across the globe.

We begin this section with definitions and then identify baselines that, if available, would enable single platform and cross-platform comparisons, strengthening observational studies. We then discuss how these baselines can inform election specific observational studies, focusing on ones that are cross platform and multi-modal in nature. Finally, we suggest different directions that are foundational for facilitating observational research about social media usage broadly and elections more specifically.

Definitions

When writing an interdisciplinary white paper, defining core concepts is an important first step. We focus on terms for which our group of interdisciplinary members had different definitions. While we recognize that researchers in other disciplines may define these terms differently, we use these definitions as our working definition for this white paper.

- **Census**: population level user measurements, population level content measurements, and/or population level network measurements.
- **Benchmark**: The state of users or content on a platform at a specific point in time. While some researchers may view it as the data used to measure the performance of a machine learning method, we use the term more broadly.
- **Random sample**: Unless it is otherwise specified, we assume the sample we are discussing to be a random sample of *users*. It is not uncommon for social media platforms to share other types of random data, including random samples of posts.
- **Baseline**: The baseline is defined as the normal state of users and/or content on a platform. This contrasts with the classic definition of baseline in an experimental study.

[11] The first draft of this section was written by Lisa Singh.

- **Replication:** The ability to use the same "environment" to repeat an experiment. In the context of observational studies, replication is confined to enabling multiple researchers to use the same observational data to validate an existing study or conduct a new study.

Observational Measurements

There are a wide variety of measurements that would be useful for election related studies. Here we identify different digital information environment baselines, beginning with overall statistics that would be useful to collect through time for different platforms and then moving to different types of samples to which it would be important for researchers to have access.

Descriptive Statistics Baselines

Time spent by people on a platform. This baseline could be used to answer questions about the amount of time people spend on different social media platforms. This not only gives researchers information about the popularity of different platforms and how different events that take place in different parts of the world, e.g. elections, protests, etc., relate to usage. Ideally, this information would be shared by a platform. However, with the presence of bots and group-level accounts, a survey sample may be more fruitful for getting specific time spent information in areas of the world having upcoming elections.

Who is on a platform. This baseline could be used to understand the demographics of people on different platforms. This would support research about the amount of time people spend on different platforms by demographic and allow for a better understanding of how the user base changes over time or as a result of specific events, e.g. Elon Musk buying Twitter, or platform algorithm/design changes. Platforms have been known to share some demographic information about their user base, but the sharing is infrequent and typically not shared by country or other spatial regions. Therefore, a survey sample or random sample with demographic inference for missing demographics are promising directions.

Languages of platform users. The multilingual and multicultural nature of platforms can be important for characterizing observations, opinions, and behaviors. Therefore understanding which languages are used on different platforms and its modality - textual or spoken, can be insightful for seeing which communities are being exposed to messaging on a platform.

User activity. Understanding activity and types of activity are necessary for knowing how engaged people and bots are on platforms. Activity can be computed using different measures of engagement, including the amount of content produced, likes on content, sharing of content, and number of followers. This will also provide researchers with a baseline for the proportion of active users on each platform versus lurkers.

In order to have a reasonable understanding of the current state, all of the mentioned baselines should be updated regularly. We do not have a good understanding of how often baselines need to be updated to make them useful or whether some baselines need to be updated more frequently, especially when major events like elections are occurring. However, in order to maintain temporal validity, starting with monthly updates for all baselines would help researchers conduct observational studies about trends and deviations; future research should determine the appropriateness of monthly updates as opposed to some other time period.

Sample Baselines

The most important types of baseline are random samples from every platform—one that is a **random set of users** and one that is a **random set of content**. Because most platforms have such long tails with respect to usage, having information about non-active users is important for understanding the overall population on each platform. It is possible to get a Census for smaller platforms already, e.g. Gab, Rumble, and Gettr. Currently, there are very few platforms for which we are able to obtain random samples. On some platforms we can obtain random samples of posts/content, but not users. This type of sampling does not allow researchers to understand the types of content that are not viewed or shared (Mneimneh et al., 2021).

Content based samples are just as important as user based samples since they improve our understanding of different types of shared content. These samples can then be used to not only understand the prevalence of different types of conversations, including political or election-related ones, but will also give us a baseline for high-quality versus low-quality content. We pause to note that researchers do not fully understand the differences between using different methods for data collection on the same platform, e.g. streaming collections vs. scraping vs. platform APIs. So keeping track of the data collection method when producing these different random samples is important for improving our methodological understanding (Trezza, 2023).

Specifically for politics and elections, the standard class of baseline topics in which we are interested are election topics discussed on a platform, election topics seen by respondents (overall and by demographic), and topics posted by respondents (overall and by demographic). Previous research has also conducted **observational studies that create clean/anonymous browsing sessions** to determine the types of content users are shown when they are new to the platform and/or have had an account for a longer period of time (Ledwich & Zaitsev, 2019). This type of study design can help us understand differences in advertising shown to different groups of users. Another important baseline is **pro-social or positive engagement** among platform users. According to platforms, creating positive engagement on their site is their overarching goal. To investigate this, we can imagine observational studies that quantify what normatively "healthy" digital media looks like, and how that differs from what platform algorithms put in front of users or what content users want/seek out. This type of study also requires understanding **the types and amount of toxicity and misinformation** platform users are exposed to and the variance of that across different platform communities. Toxicity and misinformation can be measured broadly or organized around election (or other) topics.

These types of observational studies lead to thinking more broadly about **content moderation**. How does content moderation vary by platform and by country within and across platforms? How often are content or users taken down? Measuring this is complex because content may be taken down for different reasons. A few types of "taken down" content include the following. (1) **Content that is pre-emptively moderated** by the platform, i.e. it includes keywords or images platform users cannot post. While we will never see this content, conducting algorithmic audits to see how effective different algorithms are is important. (2) **Content that is posted and then subsequently moderated** by the platform; this allows researchers to understand the lag between posting and moderation. (3) Content that **users delete themselve**s; it would be really useful to know if there is an average time span within which we expect users to delete their own content (e.g., a content creator not getting enough engagement removes a TikTok that did not go viral vs. a Twitter user deleting an account after Elon Musk bought Twitter). (4) **Content that does not get moderated or deleted**. (5) **Content that gets edited** on platforms that allow edit functionality.

Content moderation is also related to platform policies. Therefore, having an organized set of platform policies that are current will enable researchers to track the impact of different policies on deliberation and quality of conversation.

Another important group of baselines focuses on capturing **network structure**, including statistics about friends, follows, and homophily in connections. We can think about homophily across multiple dimensions (geographic; linguistic; ideology; gender; race/ethnicity, etc.). Observational research related to network structure includes questions related to the heterogeneity of networks, the amount of flow and pathways of flow for different types of political information, the existence and the formation of different types of clusters, and the role of different types of prominent individuals and accounts within the network.

Finally, while we have discussed the importance of baselines being constructed at regular **temporal intervals**, having **broad spatial coverage** is just as important. Currently, we do not have samples with adequate geographic and language representation. Benchmarking needs to be done on platforms around the world. If that is not possible, even selecting a set of countries to begin with will improve the consistency and quality of the benchmarks. Another option is to develop an international survey (or add new questions to an existing one) that contains a standard set of measures. We note that researchers have spent a lot of time developing questions about social media usage. We should borrow from those best practices.

In summary, we believe that creating these benchmarks will allow researchers to not only address the substantive meaning of theirs and others' research findings, but also improve the quality of researcher collected data. We point to two pathways. First, benchmarks will help researchers **clarify the target population** of their social media samples. Target populations are crucial for inference, but too often social media researchers rely on convenience samples without clarifying the intended population of inference. Second, benchmarks could help researchers improve the quality of their samples through **post-stratification weighting**, a technique commonly used in survey methodology to make samples more representative of a target population.

Election specifics

While all the aforementioned baselines and samples are important for understanding elections, there are some additional baselines that would be useful for elections. For example, election monitoring has clear measures for elections. It would be valuable to set up baselines that capture **global threat assessment, fraud claims, voter suppression, and polarization at the country**

level. To understand the role of political elites, a list of elites from across the world could be developed and baselines related to what the political elites share could be developed. Because of the large number of researchers studying different political elites, perhaps it would be possible to combine what researchers are already collecting to create a cross-platform baseline.

We also note that while representativeness is often the goal for election horse race type surveys, and while we care about that on social media too, we often want to understand the behavior of users at the tails or “hard to reach” populations. It is important to capture information about those that surveys may be missing.

Examples of Researcher Created Benchmarks and Large-scale Observational Studies

Here we highlight a few studies that either share benchmarks with researchers or are large observational studies that provide examples of the baselines and samples identified in the previous subsection. We note that these are just a few examples to provide a glimpse into the types of studies that can be conducted and the types of benchmarks that are available.

Different large-scale observational studies and benchmarks have been developed for Twitter/X. For example, Wolf and colleagues studied the behavior of 600,0000 Twitter users over a seven year period (Wolf et al., 2022). They observed changing behavior over time, including more posting and cohort level growth in the way Twitter is used. Hashemi and colleagues conducted an observational study about how Farsi Twitter has evolved since the Iranian Green movement in 2009 (Hashemi et al., 2022). They also released a data set of all Farsi language tweets for 500 days beginning in September 2019. Researchers have also released different types of estimates of the demographics of users on Twitter/X (Wojcik & Hughes, 2019), as well as state-level estimates (Mickey Jackson et al., 2021). During the COVID-19 pandemic, different research groups shared benchmarks to support research related to the pandemic (Tsao et al., 2021). Qazi and colleagues have also shared a large-scale benchmark for disaster incidents (Alam, Qazi, Imran, & Ofli, 2021). We also note the importance of benchmarks for specific classification tasks, including topic/conversation reach and online perception (sentiment and stance, for example). While a large number of independent efforts exist, very little benchmarking is available for multiple prediction tasks. To help mitigate this, Barbieri and colleagues created a repository that pulls together labeled Twitter/X data for different tasks (Ushio, Neves, Silva, Barbieri, & Camacho-Collados, 2022). This is an area of importance given the **large amount of textual and video information** on social media and the **small number of manually labeled benchmarks**.

While Twitter/X has the largest number of larger scale observational studies and benchmarks, other platforms also have benchmarks. For example, McGrady and colleagues shared a random sample of YouTube videos (McGrady et al., 2023). Using this sample, they estimate the total number of publicly viewable videos. The authors also hand annotated some information about the video content and processed the audio to determine the distribution of languages spoken. A number of studies have analyzed news consumption on different platforms (Pew Research Center, 2024). Because of its data availability for many years, Reddit is also a platform that has been studied more extensively. Researchers have shared fake news, toxicity, and misinformation benchmarks (Salminen et al., 2020; Salminen et al., 2020), an area of need during elections.

To the best of our knowledge, **cross-platform baselines do not currently exist**. These types of baselines are important for a number of reasons: 1) they can help better contextualize the different types of information being produced and consumed, 2) they can provide information about the impact of spread on one platform on its spread on another platform, and 3) because we are witnessing the fracturing of social media, where different users are participating on different subsets of platforms, cross-platform baselines will give us a better perspective on information consumption across a larger population. This is particularly important for cross-national understanding.

There is also a need for establishing infrastructure for comparing computational methods used for different measurements. Given the complexity of many measurements of interest, some of them are being determined computationally. However, social scientists do not have an easy way to compare the results of different methods on large-scale social media data sets or even understand which ones are the best ones to apply for specific data sets (Ladd et al., 2021). Examples of measurement that are computed using state-of-the-art algorithms include stance, misinformation, and topics. An infrastructure that has the core computation methods for a measurement within a single software package and an infrastructure to run them on benchmarks is needed to make sure of the best methods when measuring content from social media.

Conclusion

Currently, researchers collect social media data for specific studies. This is problematic because different research groups are expending energy collecting the same data instead of using a shared repository. Thanks to the survey-based efforts of Pew, YouGov, and Gallup, we increasingly have reasonable data on who in the United States participates in online digital environments and how they do so. These efforts have been conducted independently, however, with global and cross-national efforts to assess digital media participation. We advocate harmonizing these currently disjoint efforts and approaches to develop more global measures of digital media participation broadly, and election dynamics more specifically. We think this could be done by working with global surveys, like the Barometer surveys[12], to harmonize digital media questions across national contexts. This approach would require the following: (1) Liaising with organizations that run these **cross-national surveys** (like the Barometers) to investigate their current practice of measuring digital media participation and evaluate the feasibility and requirements for bringing new digital media questions into the surveys; (2) Developing a **standardized list of questions** previously validated for measuring digital media participation; (3) Developing a standardized approach for adding to these questions in subsequent survey waves to **capture the changing digital environment**; (4) Working with country experts to make sure these questions are appropriate for **country-specific contexts**. Through this type of effort, researchers would be able to better understand how digital media participation varies cross-nationally, which could in turn

[12] E.g., the Afrobarometer, Eurobarometer, etc.

inform our understanding of election dynamics in different parts of the world, and how campaigns in one region of the world inform campaigns in other regions.

Another model that may be fruitful is to make data available through a consortium of institutes and centers, given that the cost of creating these types of baselines and samples may be too high for a single institute. This institute would document the process and make data available to researchers. There would be an international committee of researchers who would meet annually to discuss the data and make strategic decisions about gaps that exist, metadata requirements, etc. Because the information environment is continually changing, the mode of collection would naturally differ from traditional survey studies. We also note that convincing platforms to work with researchers and share data is an important strategy, but relying on platforms to do so exposes researchers to the whims of the platforms. Legally mandating platforms to share data therefore also remains an important potential avenue; with the advent of the Digital Services Act (DSA) in Europe, we should begin to get our first insights into the potential of this approach.

While we have identified a number of baselines and samples of interest, the high priority ones for elections include: (1) usage of platforms by country, language, and demographics, (2) counts of humans, organizations, and bots through time, (3) activity level and follower ratios, (4) estimates of content posted per day and content taken down per day, and (5) a proportion of a stable set of election topics per day, including misinformation topics.

Finally, a great deal has been written about the ethics of using social media data for different types of studies (Hemphill et al., 2022). We advocate using **public data**, allowing for **removal from collected data sets,** and **obtaining consent** for data from private platforms like WhatsApp. Consent can be obtained from surveys, opt-in online forms, data-donation programs, and platform consent mechanisms.

Section IV: Generative AI[13]

Introduction

Developers of Generative AI systems train their models on enormous amounts of existing data and use machine learning to produce new content (expected outputs in the form of text, audio, and/or video) based on user prompts. The technology has significantly improved in the last few years, making it easier to produce increasingly realistic content. In particular, with the popularity of ChatGPT, it has become a regular part of public discourse, and researchers in many academic fields including the study of elections and the information environment are now examining its effects.

Many observers note the potential of Generative AI to produce false or manipulated content targeted at the electorate. Some recent examples from the U.S. include an audio impersonation of then candidate Joe Biden during the Democratic primaries (Seitz-Wald & Memoli, 2024), manipulated images of then candidate Donald Trump, and manipulated audio with then Vice-President Kamala Harris (Tenbarge, 2024).[14] Gen AI has also been identified as facilitating the production of content in foreign influence campaigns aimed at elections.[15] However, Gen AI can also be used in less malign ways by political campaigns, researchers, journalists, and others involved in elections. Either way, given the recent emergence of Generative AI and the nascent nature of and its use in elections, there remain many important but unanswered research questions about its use and impacts.

In this section, we review the emerging evidence and approaches to understanding the impacts of Gen AI on elections while highlighting emerging research and important gaps identified by participants at the RCN workshop. These include questions surrounding who uses these technologies in the context of elections and in what ways, what the potential impacts of Gen AI in elections are, which forms of Gen AI are most impactful in an election, what populations might be particularly susceptible to harms from Gen AI in the context of elections, and the possibility of "second-order" effects from the emergence of Gen AI.[16] There are also methodological questions including how we ought best to measure the impact of Gen AI on outcomes of interests, as well as what effects the use of these technologies have on the study of elections and information integrity more broadly. Finally, questions such as these have corollaries in the global context.[17] In the rest of this section, we examine these points in more detail.

---

[13] The first draft of this section was written by Dhanaraj Thakur.
[14] See Hany Farid (2024) for other U.S. based examples. For global examples, see https://restofworld.org/2024/elections-ai-tracker/ and https://mediaengagement.org/research/generative-artificial-intelligence-and-elections/
[15] See https://www.nytimes.com/2024/11/07/technology/russia-us-election-interference.html, https://www.justice.gov/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence.
[16] See below for more details, but at a basic level we conceive of concerns about "first-order" effects of Gen AI being related to people believing images or videos created with Gen AI to be real, whereas "second-order" effects relate to the consequences of the emergence of Gen AI on people's propensity to question whether images and videos that are in fact real could have been created with Gen AI.
[17] See for example https://mediaengagement.org/research/generative-artificial-intelligence-and-elections/.

How is Gen AI used in Elections?

One important group to consider is the ordinary voter, who might utilize Gen AI as a means to get information about elections (e.g., via a chatbot such as ChatGPT). While data is limited on the use of Gen AI generally, one recent U.S. national survey suggests that only 23% of U.S. adults have ever used ChatGPT (McClain, 2024). More importantly, the same Pew Research survey found that only 2% of Americans have actually used ChatGPT to look up information about the then upcoming Presidential election (McClain, 2024), although this may have changed as the November elections got closer. Worryingly, other researchers assessed outputs from 5 major generative AI models based on prompts about voting information and found that 51% of the results were inaccurate (Angwin et al., 2024). While the production of inaccurate voting information is worrying in and of itself, the extent to which election-related information from Gen AI models will reach or impact the public is an open research question. Generative AI can also be used by political campaigns. For example, campaigns can use Gen AI to try to **shape the image and narrative of a given candidate**. In the recent presidential elections in Indonesia, Gen AI was used by one campaign to create a cartoon-like and softer image of their candidate (the eventual winner) who was a former army general (Rayda, 2024).

Other emerging use cases include fundraising communications (Goldmacher, 2023), and micro-targeted content for voters (Simchon et al., 2024), all of which raise concurrent concerns about privacy and data protection. Campaigns can also use these tools for data analysis, to translate materials to other languages and contexts, and to connect to voters in new ways such as chatbots (Raj, 2024). The money, legal support, and technical skills to effectively use these tools suggest that currently, it will be easier for well-resourced campaigns to use these new technologies to their fullest extent (Martin et al., 2024), although the marginal payoff from having access to Generative AI capabilities may actually higher for less resourced campaigns. Consider, for example, **translating campaign materials** into multiple languages: well-resourced campaigns in the past may have had funding to do so in the past while less-resourced campaigns may not. Thus in this way, Generative AI could be a leveling force across campaigns with resource disparities.

There are also concerns that campaigns could use these tools to share mis- and disinformation about opponents, which is admittedly not a new problem. As the cost of manipulating images has fallen, we have seen examples of campaigns that have manipulated images of their opponents as part of negative attacks, which in the US can include elements of racism (Messing et al., 2016; Reed, 2022). With Gen AI there is the **added element of realism, personalized messaging, etc**., which raises new ethical questions of how political campaigns should use these tools and what kinds of guidelines and/or regulation may be needed (Martin et al., 2024).

Bad actors (domestic and foreign) have already started using Gen AI as part of new and existing disinformation election campaigns. These tools enable the spread of false, personalized, realistic content at scale. More specifically, researchers have demonstrated how OpenAI's GPT-3, an older version of ChatGPT, for example, could be used to generate text meant to discourage certain communities in the U.S. (e.g., Jewish or African-American) from voting for certain parties or from even voting at all (Buchanan et al., 2021). These models could also potentially be used to reduce costs and increase the scale of disinformation campaigns, enable real-time content generation (e.g., through chatbots), create more personalized content for audiences in different cultures and languages, and make disinformation efforts less discoverable (Goldstein et al., 2023). It is important to note that the effectiveness of a disinformation campaign, whether powered by Gen AI or not, is still dependent on other tools (e.g., social media platforms) for its spread. Thus, questions about the ability of trust and safety systems to detect and address disinformation, the degree of coordination across platforms, traditional media, and with election authorities, transparency and accountability among social media platforms, and independent researcher access to data all remain relevant.

What are potential impacts of Gen AI and how do we measure them?

Much of the previous section highlighted potential and actual use-cases. Where these cases exist, what are the actual impacts? For example, if voters do use chatbots to learn about elections and if they are exposed to false or incomplete information, to what extent does that change their behavior? To what extent do political campaigns using Gen AI find them to be effective tools? What are the differences in impact between the use of Gen AI content on social media, as opposed to non-Gen AI content, on social media when it comes to elections? How do these impacts change across different language outputs? And more generally, what about the impacts on the information environment, news coverage, trust in democratic institutions, etc.? Many of these questions are only now being addressed because the application of Gen AI tools to politics and elections is itself new.

One underlying question about impact is how persuasive content generated by these tools to target populations actually is, given the potential for realistic content and/or the human-like attributes of interaction with chatbots. Also, can these tools increase the scale and quality of persuasive content? Although more research is needed, at least one study suggests that **personalized political ads** can be more effective than non-personalized ads in the UK and that off-the-shelf Gen AI tools can be used to automatically generate these types of personalized ads (Simchon et al., 2024). A related issue is which modality of Gen AI content may be more persuasive and impactful. Although again there is insufficient research on this question, some observers (following the case of the deepfake audio of Biden mentioned earlier) suggest that audio may be the most problematic (AFP, 2024). Audio may hit the sweet spot between being complex enough to suggest authenticity in a way that text no longer can, but not so complex that it is easy enough to spot telltale signs that it was created by generative AI, the way images and video may be (at least for now).

The personalization of Gen AI tools such as chatbots also raises important questions about echo chambers where similar interactions and content may reinforce existing beliefs. For example, given the degree of personalization involved, to what extent are these tools creating unique echo chambers, and are they doing so on a larger scale than what already exists on social media platforms? If the problem of echo chambers is made worse, a related issue is the potential political bias of Gen AI models (Mowshowitz, 2024). One study found that the political preferences of a sample of Gen AI chatbots were center-left and libertarian (Rozado, 2024), suggesting that they may be more likely to reinforce those sets of political beliefs than others.

That said, the persuasive potential and personalized nature of chatbots can be used in more normatively positive ways. In one study researchers found that chatbots could help reduce beliefs in conspiracy theories (Stokel-Walker, 2024). In another, researchers demonstrated that people using an AI chat assistant when engaged in politically divisive conversations were more likely to improve the deliberative quality of their interactions with others (Argyle et al., 2023). Indeed, there is emerging work exploring how Gen AI can be used to improve online spaces for more deliberative democratic engagement (Tsai et al., 2024).

Other dimensions to consider in terms of impacts are gender and intersectionality. Deepfakes are the synthetic manipulations of identities and expressions in the form of video, images, or audio which make it appear as if someone says or does something they never did (Chesney & Citron, 2019). The majority of harmful deepfakes are pornographic in nature and are often targeted at women (Ajder et al., 2019), including politicians and journalists (Di Meco, 2019). The impact on politicians targeted by deepfakes, their staff, and campaigns can be significant. It can cause personal harm and have a chilling effect on free speech. In addition, this form of violence against women in politics shifts the narrative away from the policy contributions of those women to issues of gender, race, and image (which is precisely the goal of those behind these attacks) (Krook, 2020). Gender alone is not the only line of attack against women politicians. Intersectionality is relevant here and variables such as race, immigrant status, parental status, age, etc. should be considered. For example, one study found that women of color politicians in the U.S. are more likely to be subject to violent and sexualized forms of abuse than white women (Thakur et al., 2022), a trend that may continue with the proliferation of Generative AI.

All these potential impacts can have broader and **longer term second-order effects**. For example, the relationship between mis- and disinformation and **the decline in trust of democratic institutions** is well documented. What new concerns (if any) does the proliferation of Gen AI bring? With the diffusion of AI generated content, it can become more difficult to determine what is factual (i.e., the discussion of deepfakes earlier), pointing to the well-known problem of the “liar's dividend”, the idea that the prevalence of realistic false content enables bad actors to claim that true information is actually false (Chesney & Citron, 2019). A recent example of this occurring in an election context was when Donald Trump falsely claimed that an image from the Kamala Harris campaign of one of her rallies was generated with AI (Cha, 2024).

The potent benefits of developing methods of automatically labeling Generative AI content as such in ways that can be identified by machines and humans have received a great deal of attention; these methods are known as “content provenance” or, more colloquially, “watermarking”. However, whether these methods will actually deliver the benefits which have been ascribed to them remains an open question. One concern is that the use of content provenance methods can complicate the process through which people determine what to trust online. For example, one study found that participants conflated the presence of a provenance indicator with the authenticity of the content itself, implying that not having a watermark where users expect it could also lead to a decline in trust (Feng et al., 2023).

In addition to the important issue of trust, there is a potential relationship between the uptake of Gen AI and social and economic inequality. Consider, for example, the need for access to diverse sources of information and news for democratic participation. Gen AI tools may act as a mediator (on top of existing search engines) in how people get news and it may also influence the quality of content online (through the production of large amounts of web content by Gen AI). In these ways it can restrict access to and lower the quality of information online, particularly for those who already have fewer means of accessing information online (Capraro et al., 2023).

Methodological Concerns

Understanding these potential impacts also requires understanding the underlying makeup of the models being used. For example, what type of biases are introduced by the model and how do those influence expected outputs? The challenge here is that researchers, including those studying politics, have very little insight into how these models are developed. Machine learning has an explainability problem: it is difficult for developers to communicate how a model arrives at a particular prediction or decision given that sometimes millions or billions of interrelated parameters are involved in generating outputs. This has important implications for developing trust in these models, preventing disparate impacts across different demographic groups, and creating opportunities for redress (Shenkman et al., 2021). Efforts are being made to improve transparency in the development of Gen AI models (Bommasani et al., 2024). In addition, potential approaches to improve researcher access to data about how people actually use Gen AI tools (Nicholas, 2024) including data donations (Sanderson & Tucker, 2023) could also help researchers better understand bias and impacts.

Another issue is the impact of Gen AI on the study of elections and the information ecosystem itself. The **diffusion** of AI generated content on social media and other communication systems (e.g., private messaging groups) that political scientists and other researchers study can lead to greater noise given a **potential flood of synthetic content**. This can lead to an emphasis on the study of machine-to-machine or human-to-machine communications compared to the human-to-human interactions on which we tend to currently focus. Perhaps more concerning is the idea that we end up inundated with synthetic content produced by machines, leading to research generated by machines and reviewed by other machines, in a vicious cycle reminiscent of the "dead internet" conspiracy (Tiffany, 2021). Although this scenario is extreme given the incentives in academia to publish or perish, researchers may be pushed to take advantage of Gen AI tools to produce as much as possible. This could lead to more low quality (but mass produced) research papers in the field (Bail, 2024).

However, as is the case for other users, Gen AI also presents opportunities for researchers. It holds open the promise of **increasing the scale of research by automating analyses**. There are also new possibilities in terms of experimental research and the use of synthetic respondents (although see Munger, 2023). Large language models open up opportunities to conduct research in other languages (although this is most effective in high-resource languages, or those languages that cover most of the content online). Gen AI tools can also be used to code large scale data sets to support content analysis or other textual analytical methods (Rathje et al., 2023). Finally, Gen AI may also open up new opportunities for using large language models for classification tasks, building on many years of work prior to the release of ChatGPT and other chatbots that used fine-tuned transformer models such as BERT and RoBERTa for classification tasks (Wu et al., 2023; Terechshenko et al., 2021).

Conclusion & Global Considerations

Many of the concerns, opportunities, and open research questions mentioned earlier are also present in other political and electoral contexts outside the U.S. Indeed, some authoritarian governments are already using Gen AI to craft and spread narratives that support their regimes (Rancy, 2024) and some of the examples cited earlier show that these tools are already being used in elections globally by political campaigns and the electorate. However, it is also important to consider **relationships between countries**, and what that implies for Gen AI. For example, as with other technologies where the U.S. is dominant (most of the leading Gen AI models are developed by U.S. companies), other governments have attempted to eschew any (perceived) dependency on U.S. companies by promoting their own models or "sovereign AI" (Chander & Sun, 2023). Some governments have invested in their own large language models (e.g., India) but this could raise human rights concerns if a government can determine what words and forms of expression in a given language can be included in the model.

As noted above, foreign actors may also use Gen AI to strengthen disinformation campaigns aimed at other countries. These actors may be part of a state-level apparatus to undermine election integrity and democratic institutions in other countries, as we have observed for the last few years on social media. A recent example is that of a Russian supported network that used Gen AI to generate synthetic news media sites and social media accounts for **dissemination of that content** (Rancy, 2024). More broadly, such efforts can use Gen AI as part of more realistic phishing attacks on political campaigns, cyberattacks on elections infrastructure, or targeted harassment campaigns on elections authorities.

The proliferation of Gen AI may also exacerbate **socio-economic inequalities** across countries based on language. While large language models do offer exciting opportunities for work across languages, this is most likely the case for high-resource languages (those with more data available online such as most European languages) than low-resource languages, even if the latter (e.g., Swahili, Tamil, Bahasa Indonesia) are spoken by hundreds of millions of people (Nicholas & Bhatia, 2023). Most of the larger and popular large language models are trained on and are currently more effective in high resource languages (Wendler et al., 2024). In addition, most of the content online is disproportionately in English (for various historical/colonial and modern economic reasons). As Gen AI is used to produce more content in English (and other high resource languages) this can create a vicious cycle of exacerbating the current language inequity that exists around the world, minimizing the prevalence of certain dialects, and creating additional geo-political considerations around the proliferation of Gen AI.

Finally, another geo-political consideration is the environmental impact of Gen AI. While we have outlined some of the use cases, both positive and negative, for Gen AI in the political and electoral context, it's important to recognize that **the use of these technologies require significant resources with concomitant environmental costs**. Researchers have pointed out the significant carbon footprint (Luccioni et al., 2023) or the high levels of water consumption of Ge AI models (Li et al., 2023). It's also important to examine the potentially disproportionate impact the expansion in development and use of Ge AI infrastructure could have on low-income countries in the Majority World, which already bear the brunt of current effects of climate change.

Conclusion

The May 2024 meeting of the NYU Center for Social Media & Politics' Research Coordination Network (RCN) served as a critical step toward achieving the broader goals of our initiative: fostering interdisciplinary collaboration, advancing the study of the online information ecosystem, and building the infrastructure necessary for rigorous and impactful research on digital media and democracy. With support from the NSF's Secure and Trustworthy Cyberspace (SaT-C) program, we convened leading scholars from diverse fields to address some of the most urgent challenges posed by the evolving digital environment in an unprecedented election year.

Thanks to the dedicated contributions of our RCN members and steering committee, each of the sections in this report generated valuable insights and recommendations that are essential for moving the field forward.

Research Infrastructure: Strengthening data infrastructure emerged as a pressing and ongoing need of this field of research. Recommendations focused on fostering multinational collaborations, developing open-access datasets, establishing ethical data sharing standards, and advocating for policies that balance privacy with research needs, such as those modeled by the EU's Digital Services Act.

Experimental Research: RCN member discussions emphasized the importance of experiments in understanding individual behaviors in the digital space, particularly for sub-populations vulnerable to misinformation. Our recommendations include expanding baseline data, developing standardized tools such as browser extensions, and engaging with at-risk populations globally.

Sampling & Observational Data: Participants highlighted the need for reliable baselines in observational studies as well, to enhance our understanding of digital platforms' role in elections. Cross-platform collaboration and the creation of shared, robust baselines were identified as essential steps to provide meaningful benchmarks for election studies.

Generative AI: Researchers discussed how the rapid adoption of Generative AI impacts voter access to information, foreign influence operations, and political campaign practices. Key takeaways include the need for greater transparency in AI usage and outputs, especially regarding biases, and further research on labeling and interventions like inoculation strategies.

The outcomes of this meeting have laid a strong foundation for further research and collaboration within the RCN and in the field more broadly. To sustain momentum, we call for concerted efforts to enhance data accessibility, expand shared datasets, and build infrastructure that can accommodate the complexities of our evolving digital information environment. The insights generated here will not only inform future studies but also support the development of policies that safeguard the integrity of elections and democratic processes in the digital age.

Works Cited

Section I: Infrastructure

Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., De Vreese, C., & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research, 4*(2), 372-387. https://doi.org/10.5117/CCR2022.2.003.ARAU

Association of Internet Researchers. (2022). *Ethical guidelines 3.0*. https://aoir.org/reports/ethics3.pdf

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 830-839).

Brown, M., Sanderson, Z., Graham, S., Kim, M., Tucker, J., & Messing, S. (2024). Digital town square? Nextdoor's offline contexts and online discourse. *Journal of Quantitative Description: Digital Media, 4*.

EDMO. (2023). *EDMO releases report on researcher access to platform data*. European Digital Media Observatory. https://edmo.eu/edmo-news/edmo-releases-report-on-researcher-access-to-platform-data/

Evidence in Governance and Politics. (n.d.). *Evidence in Governance and Politics*. Retrieved October 1, 2024, from https://egap.org/

HiQ Labs, Inc. v. LinkedIn Corporation, No. 17-16783, (9th Cir. 2022). On remand from the United States Supreme Court. https://cases.justia.com/federal/appellate-courts/ca9/17-16783/17-16783-2022-04-18.pdf?ts=1650301256

IEEE. (2019, April 29). *Datasets*. IEEE DataPort. https://ieee-dataport.org/datasets

Jungherr, A. (2015). Twitter as political communication space: Publics, prominent users, and politicians. *Contributions to Political Science*, 69–106. https://doi.org/10.1007/978-3-319-20319-5_4

Lukito, J., Gursky, J., Foley, J., Yang, Y., Joseff, K., & Borah, P. (2023). ‘No reason [.][I] t/Should/Happen here’: Analyzing Flynn’s retroactive doublespeak during a QAnon event. *Political Communication, 40*(5), 576-595. https://doi.org/10.1080/10584609.2023.2241740

Methodology. The Future of Quantitative Research in Social Science. (n.d.). https://www.smrconverge.org/home/methodology/

Miller, R. (2024, March 25). Lawsuit filed by Elon Musk's X against CCDH thrown out by judge. *CNBC*. https://www.cnbc.com/2024/03/25/lawsuit-filed-by-elon-musks-x-against-ccdh-thrown-out-by-judge.html

Miller, R. (2024, May 10). Elon Musk's X loses lawsuit against Bright Data over data scraping. *CNBC*. https://www.cnbc.com/2024/05/10/elon-musks-x-loses-lawsuit-against-bright-data-over-data-scraping.html

National Internet Observatory. (n.d.). *National Internet Observatory*. Retrieved October 1, 2024, from https://nationalinternetobservatory.org/

Roeder, O. (2018, July 31). Why we're sharing 3 million Russian troll tweets. *FiveThirtyEight*. https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/

Roper Center for Public Opinion Research (n.d.). *iPoll: A comprehensive database of public opinion*. Retrieved October 1, 2024, from https://ropercenter.cornell.edu/ipoll/

Social Media Archive at ICPSR. (n.d.). *Social Media Archive*. Retrieved October 1, 2024, from https://socialmediaarchive.org/?ln=en

Lam, M. S., Pandit, A., Kalicki, C. H., Gupta, R., Sahoo, P., & Metaxa, D. (2023). Sociotechnical audits: Broadening the algorithm auditing lens to investigate targeted advertising. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW2), 1-37. https://doi.org/10.1145/3644287

Primig, F., Szabó, H. D., & Lacasa, P. (2023). Remixing war: An analysis of the reimagination of the Russian–Ukraine war on TikTok. *Frontiers in Political Science, 5*, 1085149. https://doi.org/10.3389/fpos.2023.1085149

Rathje, S., Roozenbeek, J., Van Bavel, J. J., & Van Der Linden, S. (2023). Accuracy and social motivations shape judgments of (mis)information. *Nature Human Behaviour, 7*(6), 892-903. https://doi.org/10.1038/s41562-023-01586-4

Section II: Improving Experimental Research: Baselines & Tooling

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236.

Allcott, Hunt, Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, et al. (2024). The effects of Facebook and Instagram on the 2020 election: A deactivation experiment. *Proceedings of the National Academy of Sciences* (2024). https://doi.org/10.1073/pnas.2321584121

Bail, Christopher A., D. Sunshine Hillygus, Alexander Volfovsky, Maxwell B. Allamong, Fatima Alqabandi, Diana Jordan, Graham Tierney, Tina LaChapelle, Andrew Trexler, and Austin van Loon. (2023). Do We Need a Social Media Accelerator? SocArXiv Papers. https://doi.org/10.31235/osf.io/ucfbk

Ballard, C., Goldstein, I., Mehta, P., Smothers, G., Take, K., Zhong, V., ... & McCoy, D. (2022, April). Conspiracy brokers: Understanding the monetization of YouTube conspiracy theories. In Proceedings of the ACM Web Conference 2022 (pp. 2707-2718).

Bo, H., & Galiani, S. (2021). Assessing external validity. *Research in economics*, 75(3), 274-285.

Carey, J. M., Chi, V., Flynn, D. J., Nyhan, B., & Zeitzoff, T. (2020). The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in Brazil. Science advances, 6(5), eaaw7449.

Coppock, A., Guess, A. & Ternovski, J. When Treatments are Tweets: A Network Mobilization Experiment over Twitter. Polit Behav 38, 105–128 (2016). https://doi.org/10.1007/s11109-015-9308-6

Diya, S. R. (2024, July 25). *How Bangladesh fell into an information blackout*. Tech Policy Press. https://www.techpolicy.press/how-bangladesh-fell-into-an-information-blackout/

Freedom House. (2022). *Sri Lanka: Freedom on the net 2022 country report*. https://freedomhouse.org/country/sri-lanka/freedom-net/2022

Garimella, K., & Chauchard, S. (2024). WhatsApp Explorer: A Data Donation Tool To Facilitate Research on WhatsApp. *arXiv preprint* arXiv:2404.01328.

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., ... & Tucker, J. A. - (2023a). How do social media feed algorithms affect attitudes and behavior in an election campaign?. *Science*, 381(6656), 398-404.

- (2023b) Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656), 404-408.

Javed, R. T., Shuja, M. E., Usama, M., Qadir, J., Iqbal, W., Tyson, G., ... & Garimella, K. (2020, December). A first look at COVID-19 messages on WhatsApp in Pakistan. In 2020 IEEE/ACM International conference on advances in social networks analysis and mining (ASONAM) (pp. 118-125). IEEE.

Khalaji, M. (2022, October 6). *Social Media in Iran's protests: A new public sphere?* The Washington Institute. https://www.washingtoninstitute.org/policy-analysis/social-media-irans-protests-new-public-sphere

Kleppe, M., & Otte, M. (2017). Analysing and understanding news consumption patterns by tracking online user behaviour with a multimodal research design. *Digital Scholarship in the Humanities*, *32*(suppl_2), ii158–ii170. https://doi.org/10.1093/llc/fqx030

Melo, P., Messias, J., Resende, G., Garimella, K., Almeida, J., & Benevenuto, F. (2019). WhatsApp Monitor: A Fact-Checking System for WhatsApp. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 676-677. https://doi.org/10.1609/icwsm.v13i01.3271

Möller, J., Van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? Social Science Computer Review, 38(5), 616–632. https://doi.org/https://doi.org/10.1177/0894439319828012

Mønsted, B., Sapieżyński, P., Ferrara, E., & Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using Twitter bots. PloS one, 12(9), e0184148.

Moon, W. K., Chung, M., & Jones-Jang, S. Mo. (2022). How Can We Fight Partisan Biases in the COVID-19 Pandemic? AI Source Labels on Fact-checking Messages Reduce Motivated Reasoning. *Mass Communication and Society*, 26(4), 646–670. https://doi.org/10.1080/15205436.2022.2097926

Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field Experiments on Social Media. *Current Directions in Psychological Science*, *31*(1), 69-75. https://doi.org/10.1177/09637214211054761.

Munson, S., Lee, S., & Resnick, P. (2013). Encouraging reading of diverse political viewpoints with a browser widget. In Proceedings of the international AAAI conference on web and social media (Vol. 7, No. 1, pp. 419-428).

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. Political Behavior, 39, 629-649.

Nyhan, B., Reifler, J. When Corrections Fail: The Persistence of Political Misperceptions. *Polit Behav* 32, 303–330 (2010). https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., ... & Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137-144.

Pennycook, G., Epstein, Z., Mosleh, M. et al. Shifting attention to accuracy can reduce misinformation online. Nature 592, 590–595 (2021). https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Psychological Science Accelerator (PSA). Psychological Science Accelerator. (2017). https://psysciacc.org/

Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within Google search. *Proceedings of the ACM on human-computer interaction*, 2(CSCW), 1-22.

Starbird, K. (2017, May). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1, pp. 230-239).

Törnberg, P., Valeeva, D., Uitermark, J., & Bail, C. (2023). Simulating social media using large language models to evaluate alternative news feed algorithms. arXiv preprint arXiv:2310.05984.

Wells, G., Horwitz, J., Seetharaman, Deepa. Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show." *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

Xu, Y., Chen, N., Fernandez, A., Sinno, O., & Bhasin, A. (2015, August). From infrastructure to culture: A/B testing challenges in large scale social networks. In Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2227-2236).

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtelris, N., Leontiadis, I., Sirivianos, M., ... & Blackburn, J. (2017, November). The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In Proceedings of the 2017 Internet Measurement Conference (pp. 405-417).

Section III: Sampling & Observational Data

Alam, F., Qazi, U., Imran, M., & Ofli, F. (2021). HumAID: Human-annotated disaster incidents data from Twitter with deep learning benchmarks. *arXiv*. https://arxiv.org/abs/2104.03090

Altay, S., Kleis Nielsen, R., & Fletcher, R. (2022). Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media, 2*. https://doi.org/10.51685/jqd.2022.020

Arab Barometer Survey Data. Arab Barometer. (2024, July 12). https://www.arabbarometer.org/survey-data/

Galli, A., Masciari, E., Moscato, V., & Sperlí, G. (2022). A comprehensive benchmark for fake news detection. *Journal of Intelligent Information Systems*, 59(1), 237–261. https://doi.org/10.1007/s10844-021-00646-9

Hashemi, L., Wilson, S., & Sanhueza, C. (2022). Five Hundred Days of Farsi Twitter. *Journal of Quantitative Description: Digital Media, 2*. https://doi.org/10.51685/jqd.2022.005

Hemphill, L., Schöpke-Gonzalez, A., & Panda, A. (2022). Comparative sensitivity of social media data and their acceptable use in research. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01773-w

Jackson, M., McKinstry, J., McPhee, C., Raghunathan, T., Singh, L., Traugott, M., Turakhia, C., & Nathan Wycoff, N. (2021, November). Understanding Who Uses Twitter State level estimates of those on Twitter MOSAIC Methods Brief. *Measuring Online Social Attitudes and Information Collaborative*. https://mosaic.mdi.georgetown.edu/wp-content/uploads/2021/11/MOSAIC-Methods-Brief-State-Level-Twitter-Use_11092021.pdf

Ladd, J., Ryan, R., Singh, L., Bode, L., Budak, C., Conrad, F., … Traugott, M. (2020, December 23). Measurement Considerations for Quantitative Social Science Research Using Social Media Data. https://doi.org/10.31234/osf.io/ga6nc

Ledwich, M., & Zaitsev, A. (2019). *Algorithmic extremism: Examining YouTube's rabbit hole of radicalization*. arXiv. https://arxiv.org/abs/1912.11211

McGrady, R., Zheng, K., Curran, R., Baumgartner, J., & Zuckerman, E. (2023). Dialing for videos: A random sample of YouTube. *Journal of Quantitative Description: Digital Media, 3.* https://doi.org/10.51685/jqd.2023.022

Mneimneh, Z., Pasek, J., Singh, L., Best, R., Bode, L., Bruch, E., … Wojcik, S. (2021, March 16). Data Acquisition, Sampling, and Data Preparation Considerations for Quantitative Social Science Research Using Social Media Data. https://doi.org/10.31234/osf.io/k6vyj

Pew Research Center. (2024, September 17). *Social Media and News Fact sheet*. Pew Research Center. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Salminen, J., Sengün, S., Corporan, J., Jung, S., & Jansen, B. J. (2020). Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PLOS ONE*, *15*(2). https://doi.org/10.1371/journal.pone.0228723

Trezza, D. (2023). To scrape or not to scrape, this is dilemma. The post-API scenario and implications on Digital Research. *Frontiers in Sociology*, *8*. https://doi.org/10.3389/fsoc.2023.1145038

Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. A. (2021). What social media told us in the time of COVID-19: A scoping review. *The Lancet Digital Health, 3(3)*. https://doi.org/10.1016/s2589-7500(20)30315-0

Ushio, A., Neves, L., Silva, V., Barbieri, F., & Camacho-Collados, J. (2022). Named entity recognition in Twitter: A dataset and analysis on short-term temporal shifts. *arXiv*. https://arxiv.org/abs/2210.03797

Wojcik, S., & Hughes, A. (2019, April 24). *Sizing up Twitter users*. Pew Research Center. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

Wolf, F., Lorenz-Spreen, P., & Lehmann, S. (2022). Successive cohorts of Twitter users show increasing activity and Shrinking Content Horizons. *Journal of Quantitative Description: Digital Media, 2*. https://doi.org/10.51685/jqd.2022.014

Section IV: Generative AI

AFP. (2024, February 5). *Biden robocall: Audio deepfake fuels election disinformation fears*. France 24. https://www.france24.com/en/live-news/20240205-biden-robocall-audio-deepfake-fuels-election-disinformation-fears

Ajder, H., Giorgio Patrini, Cavalli, F., & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact.* Deeptrace Labs.

Angwin, J., Nelson, A., & Palta, R. (2024). *Seeking Reliable Election Information? Don't Trust AI*. Proof News. https://www.proofnews.org/seeking-election-information-dont-trust-ai/

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., Sorensen, T., & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, *120*(41), e2311627120. https://doi.org/10.1073/pnas.2311627120

Bail, C. A. (2024). Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, *121*(21), e2314021121. https://doi.org/10.1073/pnas.2314021121

Bommasani, R., Klyman, K., Kapoor, S., Longpre, S., Xiong, B., Maslej, N., & Liang, P. (2024). *The Foundation Model Transparency Index v1.1 May 2024*. Stanford University Center for Research on Foundation Models.

Buchanan, B., Lohn, A., Musser, M., & Sedova, K. (2021). *Truth, Lies, and Automation*. Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/truth-lies-and-automation/

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J., Gigerenzer, G., Greenhow, C., Hashimoto, D., Holt-Lunstad, J., Jetten, J., Johnson, S., Kunz, W. H., Longoni, C., … Viale, R. (2023). *The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making* (SSRN Scholarly Paper 4666103). https://doi.org/10.2139/ssrn.4666103

Cha, A. E. (2024, August 11). Trump falsely accuses Harris campaign of fabricated AI crowd photos. *Washington Post*. https://www.washingtonpost.com/politics/2024/08/11/trump-falsely-accuses-harris-campaign-fabricated-ai-crowd-photos/

Chander, A., & Sun, H. (Eds.). (2023). *Data Sovereignty: From the Digital Silk Road to the Return of the State*. Oxford University Press. https://doi.org/10.1093/oso/9780197582794.001.0001

Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*, 1753.

Di Meco, L. (2019). *#ShePersisted Women, Politics, & Power in the New Media World* (pp. 1–58). The Wilson Center. https://www.she-persisted.org/s/191106-SHEPERSISTED_Final.pdf

Farid, H. (2024). *Deepfakes in the 2024 US Presidential Election*. https://farid.berkeley.edu/deepfakes2024election/

Feng, K. J. K., Ritchie, N., Blumenthal, P., Parsons, A., & Zhang, A. X. (2023). Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions. In *Proc. ACM Hum.-Comput. Interact.* (Vol. 7, Issue CSCW2, p. Article 270). Association for Computing Machinery.

Generative Artificial Intelligence and elections. Center for Media Engagement. https://mediaengagement.org/research/generative-artificial-intelligence-and-elections/

Goldmacher, S. (2023, March 28). A Campaign Aide Didn't Write That Email. A.I. Did. *The New York Times*. https://www.nytimes.com/2023/03/28/us/politics/artificial-intelligence-2024-campaigns.html

Goldstein, J. A., DiResta, R., Sastry, G., Musser, M., Gentzel, M., & Sedova, K. (2023). *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*. Stanford Internet Observatory. https://cyber.fsi.stanford.edu/io/publication/generative-language-models-and-automated-influence-operations-emerging-threats-and

Justice Department disrupts covert Russian government-sponsored foreign malign influence operation targeting audiences in the United States and elsewhere. Office of Public Affairs | Justice Department Disrupts Covert Russian Government-Sponsored Foreign Malign Influence Operation Targeting Audiences in the United States and Elsewhere | United States Department of Justice. https://www.justice.gov/archives/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence

Krook, M. L. (2020). *Violence against Women in Politics*. Oxford University Press.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models* (arXiv:2304.03271). arXiv. https://doi.org/10.48550/arXiv.2304.03271

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2023). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *Journal of Machine Learning Research*, *24*(253), 1–15.

Martin, Z., Jackson, D., Trauthig, I., & Woolley, S. (2024). *Political Machines: Understanding the Role of AI in the U.S. 2024 Elections and Beyond - Center for Media Engagement - Center for Media Engagement*. Center for Media Engagement, UT Austin. https://mediaengagement.org/research/generative-ai-elections-and-beyond/

McClain, C. (2024, March 26). Americans' use of ChatGPT is ticking up, but few trust its election information. *Pew Research Center*. https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information/

Messing, S., Jabon, M., & Plaut, E. (2016). Bias in the Flesh: Skin Complexion and Stereotype Consistency in Political Campaigns. *Public Opinion Quarterly*, *80*(1), 44–65. https://doi.org/10.1093/poq/nfv046

Myers, S. L., & Barnes, J. E. (2024, November 4). How Russia Openly Escalated Its Election Interference Efforts. *The New York TImes*. https://www.nytimes.com/2024/11/07/technology/russia-us-election-interference.html

Mowshowitz, Z. (2024, March 28). Opinion | How A.I. Chatbots Become Political. *The New York Times*. https://www.nytimes.com/interactive/2024/03/28/opinion/ai-political-bias.html

Munger, K. (2023, January 25). I strongly feel that this is an insult to life itself. https://kevinmunger.substack.com/p/i-strongly-feel-that-this-is-an-insult

Nicholas, G. (2024). *Grounding AI Policy: Towards Researcher Access to AI Usage Data*. Center for Democracy & Technology. https://cdt.org/insights/grounding-ai-policy-towards-researcher-access-to-ai-usage-data/

Nicholas, G., & Bhatia, A. (2023). *Lost in Translation: Large Language Models in Non-English Content Analysis*. Center for Democracy & Technology. https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/

Raj, S. (2024, April 18). How A.I. Tools Could Change India's Elections. *The New York Times*. https://www.nytimes.com/2024/04/18/world/asia/india-election-ai.html

Rancy, A. (2024). *Manufacturing Deceit: How Generative AI Supercharges Information Manipulation*. National Endowment for Democracy. https://www.ned.org/manufacturing-deceit-how-generative-ai-supercharges-information-manipulation/

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C., & Bavel, J. J. V. (2023). *GPT is an effective tool for multilingual psychological text analysis*. OSF. https://doi.org/10.31234/osf.io/sekf5

Rayda, N. (2024, February 4). *Indonesia Elections 2024: How AI has become a double-edged sword for candidates and election officials*. CNA.

https://www.channelnewsasia.com/asia/ai-disinformation-deepfakes-indonesia-elections-4091296

Reed, K. (2022, September 27). *Twitter Points Out Apparently Darkened Photo Of Stacey Abrams*. Yahoo News. https://www.yahoo.com/news/twitter-points-apparently-darkened-photo-150200353.html

Rozado, D. (2024). The political preferences of LLMs. *PLOS ONE*, *19*(7), e0306621. https://doi.org/10.1371/journal.pone.0306621

Seitz-Wald, A., & Memoli, M. (2024, January 22). *Fake Joe Biden robocall tells New Hampshire Democrats not to vote Tuesday*. NBC News. https://www.nbcnews.com/politics/2024-election/fake-joe-biden-robocall-tells-new-hampshire-democrats-not-vote-tuesday-rcna134984

Shenkman, C., Thakur, D., & Llansó, E. (2021). *Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis.* Center for Democracy & Technology. https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/

Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, *3*(2), pgae035. https://doi.org/10.1093/pnasnexus/pgae035

Staff, R. of W., Kim, M., Yulisman, L., & Lo, K. (2024, April 16). *Rest of world's 2024 AI elections tracker*. Rest of World. https://restofworld.org/2024/elections-ai-tracker/

Stokel-Walker, C. (2024, April 12). *Chatbots can persuade conspiracy theorists their view might be wrong*. New Scientist. https://www.newscientist.com/article/2426618-chatbots-can-persuade-conspiracy-theorists-their-view-might-be-wrong/

Tenbarge, K. (2024, August 1). *How Elon Musk and a Kamala Harris deepfake ad sparked a debate about free speech and parody*. NBC News. https://www.msn.com/en-us/news/politics/how-elon-musk-and-a-kamala-harris-deepfake-ad-sparked-a-debate-about-free-speech-and-parody/ar-BB1r2j6o

Terechshenko, Z., Linder, F., Padmakumar, V., Liu, F., Nagler, J., Tucker, J., & Bonneau, R. Comparison of Methods in Political Science Text Classification: Transfer Learning Language Models for Politics (October 20, 2020). *SSRN*. http://dx.doi.org/10.2139/ssrn.3724644

Thakur, D., Hankerson, D. L., Luria, M., Savage, S., Rodriguez, M., & Valdovinos, M. G. (2022). *An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States*. Center for Democracy & Technology. https://cdt.org/insights/an-unrepresentative-democracy-how-disinformation-and-online-abuse-hinder-women-of-color-political-candidates-in-the-united-states/

Tiffany, K. (2021, August 31). Maybe You Missed It, but the Internet 'Died' Five Years Ago. *The Atlantic*. https://www.theatlantic.com/technology/archive/2021/08/dead-internet-theory-wrong-but-feels-true/619937/

Tsai, L. L., Pentland, A., Braley, A., Chen, N., Enríquez, J. R., & Reuel, A. (2024). Generative AI for Pro-Democracy Platforms. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.5aaf489a

Tucker, J., & Sanderson, Z. (2023, November 1). *Beyond Red Teaming: Facilitating User-based Data Donation to Study Generative AI | TechPolicy.Press*. Tech Policy Press. https://techpolicy.press/beyond-red-teaming-facilitating-user-based-data-donation-to-study-generative-ai

Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). Do llamas work in English? On the latent language of multilingual transformers. *arXiv*. https://arxiv.org/abs/2402.10588

Wu, P., Nagler, J. Tucker, J., & Messing S. "Large Language Models Can Be Used to Estimate the Latent Positions of Politicians." *arXiv*, (2023): https://doi.org/10.48550/arXiv.2303.12057