



# PLATFORM RESEARCH ETHICS FOR ACADEMIC RESEARCH

Josephine Lukito

## SUMMARY

Over the past two decades, academic research on digital platforms — such as social media, websites, blog posts, and digitized content — has proliferated. But how do we know if these studies are conducted ethically? And what does it mean to conduct “ethical research” in the context of studying digital platforms?

This report lays out the state of current platform studies ethics, the challenges of building ethical frameworks for this type of research, and potential solutions as proposed by researchers studying digital platforms and research ethics. Interviews with academic researchers emphasize the need for building consensus, ideally through coalitions, and for supporting research infrastructure that prioritizes clear and transparent ethical practices.

What is clear from the findings is that researchers, platform users, companies, politicians, and funders must work together to support ethical research practices that are flexible and yet guided by the shared principles of research for the public and minimizing user harm.

### SUGGESTED CITATION:

Lukito, J. (April, 2024). Platform research ethics for academic research. Center for Media Engagement. <https://mediaengagement.org/research/platform-research-ethics>

# TABLE OF CONTENT

▶▶ **01 EXECUTIVE SUMMARY**

▶▶ **02 TABLE OF CONTENT**

▶▶ **03 KEY FINDINGS**

▶▶ **04 INTRODUCTION**

Understanding Platform Research Ethics **05**

Stakeholders **06**

Research Design **07**

▶▶ **08 ETHICS IN CURRENT RESEARCH DESIGNS**

Understanding Research Design **08**

Independent vs. Collaborative Research **10**

Privacy-Oriented and Public-Oriented Ethics **13**

Balancing Research Principles **14**

▶▶ **15 CHALLENGES TO BUILDING ETHICAL FRAMEWORKS**

Lack of Consensus around Ethical Procedure **15**

Inconsistent or Weak Ethical Review **17**

Limited Infrastructure **19**

▶▶ **20 BUILDING ETHICAL STRATEGIES FOR ACADEMICS**

Conduct Ethics Research **22**

Coalition Building **23**

Research Infrastructure as Ethical Leadership **25**

▶▶ **27 CONCLUSION AND RECOMMENDATIONS**

▶▶ **29 METHODOLOGY**

▶▶ **30 ACKNOWLEDGMENTS**

# KEY FINDINGS

## Challenges

Three key challenges have the potential to threaten the ability to do academic research on digital platforms, let alone do that research well:

### A Lack of Consensus Around Ethical Procedures

- This is one of the most consistently mentioned challenges, particularly when concerned with privacy efforts.
- Procedures for efforts such as informed consent and anonymization need to see some form of consensus.

### Inconsistent or Weak Ethical Reviews

- There is concern about reliance on Institutional Review Boards (IRBs), which vary in the advice they give and often treat platform research as exempt from review.
- Some researchers are seeking alternative ways to conduct ethical reviews, such as reaching out to research ethicists.
- There is a need for researchers to have more comprehensive training on ethics, particularly if they must make their own decisions regarding ethical research practices.

### Limited Infrastructure

- Limited platform research infrastructure makes it more difficult to build consensus around ethical procedures and review.
- Funding, which tends to be project-by-project, may contribute to this issue. For sustainable infrastructure that supports a field of research, longer-term funding is necessary.
- Researchers are eager for feedback, support, or recommendations for how to improve both their ethical framework and ethical practices.

## Recommendations

To build consensus around platform research ethics, changes are necessary in three main areas:

### Ethics Research

- More support for ethics research is essential, particularly for conducting ethical review and developing ethics procedures.
- Professional academic organizations should recognize and reward ethics research in their field.
- Professional academic organizations should adopt ethical guidelines.

## **Coalition Building**

- Research associations should develop ethical guidelines for researchers within their discipline.
- Researchers should organize in-person convenings dedicated to platform research ethics.

## **Research Infrastructure**

- There should be a platform research or digital data ethical review that operates within or in tandem with Institutional Review Boards.
- Funding for research infrastructure should ask about the collaborative or independent nature of the research, as well as whether the infrastructure uses a privacy-oriented, public-oriented, or hybrid approach to research ethics.
- There should be funding for long-term infrastructure, including, but not limited to, tools for anonymization, data archives, and benchmark datasets.
- Researchers should engage with policymakers to develop a legal framework for platform research that can govern data access and set boundaries regarding what researchers can do with the data.

# **INTRODUCTION**

Several years ago, I was presented with the opportunity to study public and private digital platform content that had been gathered through non-traditional collection methods. I was both curious and concerned about the data – even though I did not collect it myself, I felt as though using it would implicitly advocate for these types of data collection strategies.

Given these challenges, I reached out to a U.S. Institutional Review Board (IRB), an administrative ethical review body that is often responsible for ensuring that no harm is done to human subjects during academic research. I expressed my concerns, noting that while there was some public value to the data, I was concerned about how the data had been collected. I was encouraged to complete an IRB proposal, wherein I noted my own ethical qualms about using the data.

Two weeks later, the study was approved and classified as “exempt,” with no follow-up regarding my concerns about the data.

While I ultimately chose not to pursue the project, the experience was eye-opening. Could I be sure that I was conducting my work ethically? And, without an ethical review, what are the risks for researchers like me (and their institutions)? Over time, I became increasingly concerned about the intended and unintended consequences of my work.

For better or worse, in trying to understand and develop an ethical framework for my research, I realized that I was not alone: The challenge of understanding platform research ethics is one that many internet researchers have struggled with.

Importantly, this is not a report about what ethical choices researchers should make. Such choices are highly dependent on context and method, and one unified ethical framework would likely not work for such a varied field.<sup>1</sup> Instead, this report lays out the challenges for academics navigating research ethics and identifies places where consensus can be built. To access additional resources on platform research ethics, or to submit your own resource, please visit our [working document](#). Support for this report was provided by NetGain Partnership.

## **Understanding Platform Research Ethics**

“Platform research ethics” refers to a framework for planning, conducting, and reporting research on digital platforms in such a way as to benefit a society and its citizens. Platform research ethics are normative – they reflect what the researcher considers appropriate or inappropriate research practice. Despite the growth of research on digital platforms, ethical guidelines have not been updated to consider the risks or challenges of such research.

To explore the concept, we break the phrase down into its three parts.

### **Platforms**

Platforms typically refer to digital and mediated spaces online that facilitate some sort of interaction between users. This interaction can be communicative, such as in a forum, but it could also be economic, such as a transaction between a seller and a buyer.<sup>2</sup> Platforms can also be scoped out more narrowly, focusing on social media platforms as a unique type of platform. This report takes the approach of the latter, but the findings can also apply to other types of platforms.

Platforms are increasingly critical to the function of societies. We use platforms to communicate with each other, to exchange information, to make financial transactions, and to manage our day-to-day lives. Naturally, this has raised questions about whether people’s dependence on digital platforms is ultimately beneficial or harmful.

### **Research**

If research is the empirical study of a natural or social phenomenon, platform research refers to the empirical study of platforms (either one, several, or the constellation of platforms that contribute to the digital information environment).<sup>3</sup> This includes, but is not limited to, research about platforms’ moderation techniques,<sup>4</sup> the relationship between platforms and news organizations,<sup>5</sup> and the effects of digital platforms on individuals.<sup>6</sup>

Because many academic fields are interested in the development and effect of this technology, platform research is an interdisciplinary smorgasbord that includes long-standing fields in STEM, the social sciences, and humanities, as well as comparatively newer scholarly areas such as human-computer interaction and science and technology studies.

## **Ethics**

In the broadest sense, ethics refers to a moral framework or set of principles that guide people's perception of "right" or "wrong" behavior. Research ethics, more specifically, is often defined as "the ethics of the planning, conduct and reporting of research."<sup>7</sup> Just as people vary in their ethical framework, researchers too vary in their ethical approaches, even among researchers using similar data and methods.<sup>8</sup>

While ethical frameworks can vary between researchers (and even for one researcher between two projects), institutional efforts to mitigate risk have resulted in procedures that systematize ethics into a procedure or a set of steps to complete. This approach to ethics is not necessarily bad – for a topic as subjective as ethics, it is still necessary to have a proverbial line in the sand to determine what is or is not acceptable in ethical research. However, research ethics thus far has focused on medical research,<sup>9</sup> rather than more contemporary methods,<sup>10</sup> which has resulted in few ethical procedures that are contextualized for platform or big data research.

Additionally, research ethics policies tend to be developed retroactively rather than proactively, meaning that few ethical guidelines are developed until a clear ethical violation has occurred. An important example of this is the Belmont Report, a report on research ethics by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. This report was a direct outcome of the Tuskegee Syphilis Study, during which the U.S. public healthcare service withheld care for 400 Black men with syphilis. This example highlights the critical need to develop guidelines before harm is done.

A key challenge for platform research ethics is the divergence between legal policy, professional ethics norms, and what can be accomplished in practice. Simply put, the novel ways in which platform research introduces risk have not been adequately addressed. This creates challenges for academic researchers who seek to conduct their research ethically but do not necessarily know how.

## **Stakeholders**

Before discussing platform research ethics, it is important to acknowledge the relevant stakeholders. These actors sometimes work harmoniously but are sometimes in conflict with one another. While there are other relevant stakeholders (e.g., non-academic collaborators, universities, microworkers, and third-party companies), discussion of these stakeholders is beyond the scope of this report.

## **Academics and Academic Staff**

This report focuses foremost on academics and academic staff. Academic researchers often have considerable and unique access to social media data and largely have independent control over what they choose to study. While there is certainly important research that is conducted outside of the academy, academic researchers must adhere to specific professional norms and institutionalized ethical review processes in a way that non-academic researchers do not. For example, in the United States, academics are subject to Institutional Review Boards, while researchers within platforms may voluntarily have their work reviewed by an IRB but are by no means required to do so.

## **Platform Users**

Platform users are often the individuals being studied, both as individuals (such as through data donations and survey work) and as groups and communities or social networks. Depending on the users studied, academics can sometimes be in contention with or in collaboration with the users they are studying. For example, academics can work with users to develop community-based participatory research practices.<sup>11</sup> The study of some online communities (such as pro-violence extremists), however, can also invite harassment.<sup>12</sup>

## **Platform Companies**

Most platforms have policies and terms of service related to the use of their data for both users and researchers. Platforms will often make their data available (to researchers, third-party creators, archivists, and other groups) through application programming interfaces (APIs). However, API access exists at the whims of the platform.<sup>13</sup> As a result, some platforms (such as [Twitter](#)) can unilaterally revoke access to researchers, moderators, and content creators leaving them at the mercy of a platform's whims.

## **Policymakers and Political Actors**

Politicians and political actors are also important stakeholders because they have the capacity to regulate platforms and determine what obligations platforms have to provide data for research. For example, the [Digital Services Act](#) may facilitate data access for European researchers. However, politicians have also banned access to specific platforms,<sup>14</sup> making it more challenging to study their content.

## **Foundations and Funders**

Foundations and funders have the capacity to bridge networks of different stakeholders for a more transparent, safe, and open information environment. Funders can also play a critical role in encouraging researchers to reflect on the ethical practices of their funded work.

## Research Design

To understand how researchers consider platform research ethics, a series of interviews were conducted with 18 academic researchers, ranging from research staff to endowed professors and center directors. The discussions covered the interviewees' research, the ethical and practical challenges they faced, and their recommendations for the future of platform research. Owing to the sensitive nature of this subject, participants have been anonymized. Additional information, including interviewees' field and rank, can be found in the Methodology section at the conclusion of this report.

Given the highly interdisciplinary nature of the subject, researchers from a variety of different fields were intentionally sought. Of the 18 interviewees, six were computer scientists or engineers, five were communication scholars, three were information scholars (primarily in the areas of human-computer interaction or science and technology studies), two were psychologists, two were political scientists, and one was an economist. It is worth recognizing that nearly all the interviewees engaged with multiple disciplines, as evidenced by their multi-disciplinary conference attendance, publication in interdisciplinary and general journals, joint appointments (primarily in communication, political science/government, and information schools), and degrees (six participants held degrees in a discipline that is different from their current department's).

## ETHICS IN CURRENT RESEARCH DESIGNS

Research ethics informs research design and vice versa. This section summarizes how researchers have studied platforms and identifies three themes: (1) independent and collaborative research, (2) privacy- and public-oriented research ethics, and (3) balancing research decisions.

### Understanding Research Design

Broadly speaking, there are three stages to conducting platform research: data collection, analysis, and presentation.<sup>15</sup> At each stage, scholars rely on a range of methods. For example, researchers use different approaches for collecting data, including both traditional approaches, such as interviews or experiments, and computational approaches, such as APIs, data donations, and scraping tools. Data collection may be the most ethically fraught stage of platform research, as researchers must make decisions about their collection approach, how much data they are gathering, and how intrusive their research design may be.<sup>16</sup> Through this process, researchers must decide how to collect, process, analyze, and present the data. As one researcher in the communication field explained, "There are a million little decisions that happen between the start and the end of a project."



Once researchers have amassed this data, they apply a variety of analytical strategies to make sense of it, including, but not limited to, computational methods, traditional quantitative methods (e.g., experiments and surveys), as well as qualitative methods (e.g., digital ethnographies and interviews). Social media data can also be combined with other datasets,<sup>17</sup> helping scholars blend methods and approaches. Once these analyses are complete, researchers write up their results and, if they can, share the data and code used for the analysis. While written chronologically, researchers often iterate between these steps to refine their study.<sup>18</sup> For example, if a researcher analyzing social media data realizes that their collection is incomplete, they may re-collect new data with a revised query.

The perceived ease by which social media data can be collected and analyzed was originally viewed optimistically<sup>19</sup>— researchers saw social media data as a vast treasure trove of information, which is true to some extent. However, the bright potential of social media research may have also blinded us to a stagnant ethical framework that has not kept up with the rapid pace of platform research.

Here, social media research ethics refers to the value framework that researchers use to ensure their studies on social media data are done responsibly — inclusive of minimizing harm, being transparent about research procedures, and respecting the confidentiality of participants, among other things. Social media research ethics is both an artifact of past research ethics frameworks and a forward-thinking approach seeking to guide researchers toward producing trustworthy social media research. As a result, ethical frameworks vary as much as the methods that researchers use to study platforms.<sup>20</sup>

A lack of a consensus around research ethics does not mean that researcher ethics does not matter to researchers. On the contrary, the lack of a consensus drives researchers to worry considerably about whether their work is ethically sound. As one researcher working in a school of information noted, “There’s incredible recognition across the broader research community that these have long been wholly inadequate to addressing core ethical concerns for those of us who are working with social media, digital data more broadly.”

Sharing similar concerns, a computer scientist highlighted the importance of ethics for the advancement of platform research, “we really need consensus because this empirical inability to have reasonable ethics, ethical decision making is damaging not only in the credibility of our field, but its validity.”

However, researchers also noted that building consensus around ethics has been challenging, especially because researchers across (and within) fields can have very different opinions on how data should be collected, handled, and shared. One communication researcher explained this succinctly: “When you do interdisciplinary research, people don’t necessarily come to the table with the same ethical frameworks.” Several of the interviewees noted disagreements regarding how to handle both data

collection and data sharing practices. For example, one computer scientist focusing specifically on digital privacy highlighted past struggles collaborating with “traditional kinds of computer scientists [who] just don’t see things in the same way because they just see it as data. [They think] this is a network, we grab data, why is this an issue?” Tensions around ethics, therefore, can create problems for essential interdisciplinary collaborations.

Cumulatively, these comments suggest that platform research ethics is at a standstill: Researchers care about ethics but are concerned that disagreements hinder the ability to build consensus around specific recommendations. This highlights how ethics continues to be a highly individualized approach, particularly given that the ethical frameworks of the past may not be able to account for the unique or exacerbated challenges of social media research.<sup>21</sup>

We next explore two dimensions of platform research for academics: independent vs. collaborative approaches, and privacy-oriented and public-oriented ethical frameworks. Both of these dimensions highlight the varied and complex ways that researchers study platform data.

## **Independent vs. Collaborative Research**

One factor contributing to ethical decision-making is the research design’s reliance on platform access or collaboration. Research projects fall along a spectrum of research designs, with collaborative research as one pole and independent research as the other.

Collaborative research approaches rely on cooperation with platforms. One recent example of this is the 2020 Election Research Project,<sup>22</sup> where academic researchers and Meta researchers collaborated on a series of experiments. Cooperative approaches have the advantage of providing potentially unique access to data and updates. Scholars also noted that collaborative research could synthesize ethical approaches between academic and industry researchers. A computer scientist who worked with industry researchers said, “There are some of us who consult for industry. These social platforms do internal research, they want to study something internally. It’s ethically challenging. So they bring folks like us in to consult, to give them some framework so they can make a decision or justify it up the chain.”

However, this research relies heavily on permission from the platform,<sup>23</sup> which may not always be provided (and, in fact, even if a platform provides data at one point in time, there is no guarantee that this access will be permanent or reliable). If a platform is highly consequential but does not permit researchers to examine even their public data, what is a researcher to do? This is particularly problematic with regard to data access because this data may not be accessible without the platform’s consent.<sup>24</sup>

Aside from providing data access, platforms can also impose stipulations on what a researcher studies and publishes, from demanding advanced review of a study to halting the publication of a finding that puts the platform in a bad light.<sup>25</sup> Platforms can be selective with whom they collaborate with, potentially shutting out researchers in smaller institutions or with fewer resources.

The other side of this spectrum is independent research, which does not rely on platform permission to conduct research. Independent research methods are popular when scholars are performing research that is critical of digital platforms;<sup>26</sup> for example, when studying the harms induced by social media consumption.

Scholars conducting independent research often seek permission directly from users. For example, researchers working on data donations research and research relying on browser extensions often obtain informed consent.<sup>27</sup> This process, which circumvents the need for platform-provided permission, has gained popularity in social science disciplines. One communication researcher highlighted the benefits of combining user-provided data with other data, such as surveys:

“

We call this a user-centric view of analyzing [platform] data. What we have is a panel of users that have donated their data. They allow us, with full informed consent, access to their [platform] timeline data. In addition, we have survey data that we collected from them, so we can connect the topics they’re actually interested in with their [platform] timeline.

Another type of independent research is conducted using publicly accessible content, regardless of whether the platform has given permission to collect that data. The most common way to access this data is through web scraping, a process for the automated collection of webpages (this is also sometimes referred to as web crawling or web extraction).<sup>28</sup> Several interviewees highlighted the advantages of scraping, not only to study individuals and the content they produce but also to gather information about projects and code from websites such as GitHub (a developer platform often used to share code between programmers). However, other interviewees, such as one from an information school, also mentioned that scraping was a “last resort” strategy when other collection methods were not possible: “I’m generally willing to use an API if there’s one available, but sometimes the information that’s provided in the API is not the information we need. I don’t mind scraping, but I try to do so in a way that doesn’t put extra pressure on the servers of the platform I’m scraping.”

Because independent research is conducted separately from the whims of the platform, this research is seen as more risky, both legally and ethically. Even when informed consent is provided by participants, platforms still have the ability to [block research](#) that relies on browser extensions or apps. The lack of guidelines around acceptable and unacceptable research practices also makes researchers (academic or not) vulnerable to lawsuits,<sup>29</sup> potentially chilling important platform research. Whereas research relying on platform collaboration may fall back on the platform's permission to justify its ethical framework, independent research is expected to use an ethical framework that still protects the users but can be at odds with the platform.

It is important to note that several researchers employ both strategies, choosing to be more collaborative or independent depending on the project. For example, one political scientist's body of work applied a range of approaches, with some collaborative projects and other independent ones. Acknowledging the advantages and disadvantages of both strategies, the researcher also said:

“When we don't work in collaboration with the platforms, we're always at risk of something happening with the platforms whether it's adversarial, or just a totally orthogonal change ... And when you work with the platforms, you have to be incredibly careful. You're at the whim of the platforms in other ways and you have to think through how you try to preserve the integrity of the research.”

Additionally, some approaches are not purely collaborative or purely independent and are, instead, a hybrid of the two. For example, researcher-specific API access is collaborative in the sense that it relies on platform permissions for access, but is flexible enough that researchers can conduct research on the data with minimal platform constraints. Researchers may use this hybrid, semi-independent approach until, hypothetically, ownership of the platform changes, causing researchers to seek out alternative, independent approaches when the researcher API closes.

Because platforms themselves tend to be opaque in their data collection and sharing policies, some researchers are skeptical about platform permission as equivalent to either an empirically justifiable or ethical approach.<sup>30</sup> For example, one researcher in the computer engineering field noted, “Something that we're struggling with right now is platform transparency ... I can't see any manner of increased transparency in the near future, and whole research teams are gone at [social media platform].” A lack of transparency not only decreases researchers' trust in the data but also creates ambiguity around whether the data are being shared ethically.

A computer scientist also explained that platforms and researchers are motivated by different goals, making collaborative research about the societal impact of that platform difficult, “Back in the day, we thought, ‘Oh maybe we will be able to work with these companies and they’ll give us some data.’ I did an internship at [social media company] and I realized pretty quickly that this was not going to work out.” The researcher suggested that this may be for two reasons: user privacy and a concern for providing data that will make the platform look bad.

Researchers also noted that platforms themselves tend to hide behind vague terms of service that ultimately limit, rather than expand, research. As one researcher at an information school argued, “Terms of service can be problematic and I feel that it doesn’t allow for independent research. [Social media] is basically controlling who can research their platform and what questions they can ask using their API data.” This aligns with the literature on terms of service; for example, a review of over 100 social media terms of services finds that these documents tend to be both vague and overly broad – after all, these documents are meant to protect the company, not to protect users or enable robust research.<sup>31</sup>

As others have noted,<sup>32</sup> we cannot conflate research ethics and platform permissions: to collaborate with a platform is not inherently ethical or unethical. Likewise, independent research – even research that violates terms of service – is not inherently ethical or unethical.

## **Privacy-Oriented and Public-Oriented Ethics**

Aside from a researcher’s relationship with a platform, another factor that may impact research ethics decisions is whether a researcher treats data as more private or more public. Researchers practicing privacy-oriented ethics are most concerned with user privacy and tend to emphasize informed consent, anonymization, and the right to be forgotten. Researchers working with data that are perceived as more sensitive in nature (e.g., data with sensitive information such as government identification numbers or private messages) tend to take a privacy-oriented approach.

By contrast, researchers practicing public-oriented ethics tend to argue that public-facing data should be permissible to study. Practitioners of public-oriented ethics are most interested in conducting research on data that are in service of the public good.<sup>33</sup> One researcher at an information school highlighted the overemphasis on individuals and underemphasis on societal considerations, “We tend to focus on individual risks, whether we think about them or ignore them, but we’re ‘hand-wavy’ about social benefits.” Practitioners of public-oriented ethics also tend to advocate for more open science practices that allow for data sharing, which they argue democratizes platform research.<sup>34</sup>

A researcher who identifies as a public-oriented researcher and teaches in the field of economics emphasized the public nature of the data in their justification of its collection:



I'm probably on the more liberal interpretation of what people are agreeing to when they post things publicly ... I think that if you post something publicly, it's there and it's there forever. And the fact that you deleted it doesn't really change anything at all in terms of my moral or ethical obligations in its use.

Another researcher teaching in the communication field described this as a maximalist and minimalist approach, where maximalists are “people who say you could never use Twitter data unless you get consent from all those people who tweeted” (i.e., privacy-oriented research ethics). By contrast, minimalists use and share public data, with fewer qualms about directly quoting users.

Given other scholarly pressures – including the trend towards open science practices – these ethical positions are often placed at odds with one another. One political scientist explained this conflict succinctly, “There is an inherent tension between the need for data access to build knowledge on the one hand, and the need to respect privacy and data protection, principles and best practices on the other.” The tension between these ethical positions is not new, as “the root of the challenge [between these two perspectives] is whether platform users understand that their data is being used for research.”<sup>35</sup>

But, realistically, and similar to the independent-collaborative spectrum, privacy- and public-oriented research ethics fall along a spectrum. For example, a researcher may rely on public-oriented ethics in the data collection (e.g., fine with scraping large amounts of public data) but use more privacy-oriented ethics in data sharing (e.g., will not share data openly), particularly as anonymization efforts can be imperfect.<sup>36</sup>

## Balancing Research Principles

While researchers, regardless of their methodological approach or ethical framework, agreed that ethics are motivated by protecting “people,” both as individuals and as members of society, they also emphasized the challenge of balancing multiple research principles, including ethical principles (e.g., do no harm, serve the public good) as well as research principles (e.g., sharing datasets, in the spirit of open science, that are [findable, accessible, interoperable, and reusable](#)).

Balancing these considerations can give the impression that researchers must be experts in platform ethics for each project they do. One researcher teaching in the computer science field noted that this can be daunting for researchers, “In order to do ethics research you

have to be trained in some ethical approach. And that’s an intimidating background. There’s a 2,000-year history or 4,000-year history of ethics ... I think it’s intimidating for people. It’s intimidating for me.” Another researcher said that this framing leads to the perception that ethics is this vague yet tangible silver-bullet concept, “[Ethics] is just thrown in the air as this ‘hoity-toity’ [concept]. Oh, you should think about ethics.”

But some research ethicists argued that being ethically informed is important for all researchers, “I do think there’s a need for people who are not making their career on ethics to be ethically informed ... To be ethically informed does not mean reading a massive body of literature that the full-time ethicists have created. It means reading a wiki page and downloading a tool on GitHub, and applying it to your research.”

Unfortunately, with limited ethical guidelines regarding how academic researchers should treat platform data, researchers are forced to develop highly individualized and varied approaches. Some do not conduct additional ethical reviews beyond IRB. But others, increasingly, are seeking out guidance from far more than their review board, including from legal counsel (both within and outside of their university), from ethics researchers, from funders, and from professional organizations.

## **CHALLENGES TO BUILDING ETHICAL FRAMEWORKS**

Platform research ethics has struggled to build a shared ethical framework. One reason for this may be the misconception that there is a unified way to distinguish “ethical” from “unethical” research. In reality, there likely will never be a universal consensus about research ethics,<sup>37</sup> especially given that different disciplines have different approaches to ethics. However, the limited legal and procedural guidelines surrounding platform research ethics have also caused researchers to seek out guidance from other sources (e.g., professional organizations and lawyers), many of which may give conflicting advice or not conduct an ethical review at all. As a result, researchers are left to make ethical determinations themselves, with few resources on how to determine a project’s ethical risks.

Highlighted below are three challenges facing research ethics for platform studies: a lack of consensus around ethical procedure, inconsistent or weak ethical review, and limited ethical infrastructure.

### **Lack of Consensus around Ethical Procedure**

One of the most consistently mentioned challenges (both by interviewees and in the literature) is a lack of consensus around ethical procedure,<sup>38</sup> particularly around privacy efforts. This is related to, but different from, a lack of consensus around ethics.



For example, participants disagreed regarding whether they thought informed consent was necessary. Projects about people’s platform information feed will often take a privacy-oriented approach, resulting in a greater emphasis on informed consent. However, projects involving public figures are less likely to consider informed consent. This is a challenge of ethical decision-making: is informed consent essential to consider platform research as “ethical”?

Even for researchers using publicly accessible data or “benchmark” datasets (popular datasets that are used within a field, often to compare computational tools), there was disagreement regarding the necessity of informed consent. For example, one computer scientist lamented, “I didn’t realize until recently that I don’t even know whether people consented to being part of a benchmark dataset.” Continuing their line of thinking, the researcher also noted that a lack of informed consent is especially problematic for visual content: “People are now becoming part of a benchmark dataset about facial features, without knowing it. These people don’t even realize it.”

However, when it comes to the *procedure* of informed consent (i.e., how researchers should request informed consent), both privacy-oriented and public-oriented researchers acknowledged that consensus was necessary. In other words, *if* researchers wanted to get informed consent, there should be agreement about *how* researchers should request informed consent. While ethics reviews have some standards and routines regarding how to solicit informed consent, the applicability of these procedures to platform research is unclear. For example, how would one go about soliciting informed consent from thousands or millions of social media users? For projects relying on web-tracking data, researchers remain unclear about how frequently they need to solicit informed consent, or whether traditional informed consent forms should be modified to account for the collection of digital data. As one political scientist succinctly explained, “Even when my colleagues who use web-tracking data go through the process of IRB, there’s the question of, ‘what does informed consent look like?’”

Another area with procedural ambiguity is anonymization. Generally, anonymization is understood to be the removal of personally identifiable information in a dataset.<sup>39</sup> However, this practice is much easier said than done. As one researcher at an information school noted, “Data anonymization is problematic because we see a lot of overly simplistic interpretations.” The same researcher elaborated on this remark by using differential privacy as an exemplary attempt at anonymization, “Differential privacy is something people throw around all the time, but it’s been over a decade, and we’re still struggling with differential privacy, so I don’t think that it’s the savior we originally talked about it being.”

One political scientist highlighted the challenges of completely removing personally identifiable information. They argued that for their method, “there’s just no way for it to not



be personally identifying because we are collecting so much data.” Part of this, notably, is because even in a space that is accessible to anyone else on the internet, users are likely to disclose personally identifiable information.

Some researchers balance this challenge by limiting what they collect or by anonymizing to the best of their ability. One researcher teaching in the computer science field said, “We do try to fuzz [the data] slightly so that we can adequately protect their privacy. If you’re smart enough, you could figure out who it is. But my goal is to give them at least a plausible deniability so someone is not like, ‘I’m going to go Google these people and look everything up.’”

## **Inconsistent or Weak Ethical Review**

Another key challenge of platform research ethics is weak ethical training and review. U.S. academics and researchers who work with U.S. academics rely primarily on Institutional Review Boards (IRBs), administrative groups that are responsible for reviewing and approving human subject research.<sup>40</sup> IRBs have been the primary ethical gatekeepers for academic research, deriving their authority from the Office for Human Research Protections in the Department of Human Health and Services and from the Belmont Report. The demand for an ethical review process stemmed largely from human-subject research atrocities before and during the 1970s, including the infamous “Tuskegee Experiment.” Since then, there have been few changes to the Office of Human Research Protections’ policies, despite the evolution of society and research over the past half-century.

Needless to say, platform research ethics is hardly a consideration in this framework. As one participant explained, “When the Belmont Report was written, which is what our current regulations are based on, there was no such thing as social media. And so our current guidelines are very permissive about data reuse.”

And, because of the piecemeal nature of IRBs (each university has its own institutional review board), researchers have expressed a wide array of opinions about IRBs they have worked with (of the 18 interviewees, 14 mentioned institutional review boards). As a researcher at an information school described:



There are some IRBs that are very good. And they can give good advice. And then there are some IRBs who are not there yet, who can’t give good advice, who haven’t really thought about these issues, and don’t have a lot of experience ... It’s frustrating for researchers because they sometimes get really good advice, and they sometimes get not helpful advice, and [researchers] don’t know what to expect in advance.

Interviewees also raised concerns about how IRBs treated platform research as exempt from review or “not human subject” research. One computational social scientist even described it as a waste of time:



I’ve actually stopped just taking my proposals to the IRB because it’s a huge waste of my time because the data is public and I don’t get any private information, and people don’t interact with us. The IRB says, ‘this isn’t our problem.’ But, you can still have ethical consequences that impact people even when IRBs have no oversight.

Some researchers also made selective decisions about when to reach out to an IRB. For example, one researcher working in computer science explained, “We have guidelines for when we would need to involve an IRB ... For 90% of our work, we have not needed to go to an IRB because we’re observing public data.”

One researcher working in the field of communication who did submit regularly to IRB also brought up how their research is considered exempt because it is “public data,” “I always go to IRB, but it’s almost always exempt because it’s public.” Because IRBs tend to treat public platform data as (by default) exempt from review, researchers end up treating the IRB process as simply something to do rather than as a useful ethics practice. This is highly problematic, as even public data can, and often do, have personal identifiers.<sup>41</sup>

Rightfully, researchers also raised concerns that IRBs were not so much intended for ethical review as they were useful for curtailing risks. Said one researcher, “IRB is not really specifically about ethics. It’s about trying to minimize a particular set of harms.” In other words, because IRBs were born out of a reaction to unethical research involving human bodies, IRBs have been less suited to handle other forms of harm to individuals.

These comments and experiences, which align with previous literature and writing on IRBs,<sup>42</sup> highlight two things. First, researchers view IRBs as insufficient for ethical review. While the process certainly makes conducting platform research easier, some interviewees also expressed discomfort with the ease of the process, “[Researchers] say the [platform] data is public and IRB said it’s exempt and we move on. And it was too easy, too quick.”

Second, researchers are seeking out alternative ways to conduct ethical reviews that go beyond submitting for IRB approval. For example, some academics reached out to research ethicists as a sort of external review for large projects. Several researchers also indicated having internal policies for anonymization or removing data when requested. One ethicist lauded these efforts, “[Researchers] say, ‘Well, the IRB said this is not human subject data, but we went and did X, Y, and Z, even though IRB was not giving us any guidance,’ and that’s amazing.”

However, interviewees also expressed a need for researchers to have more comprehensive training on ethics, particularly if they must make their own ethical decisions. One political scientist noted, “Ethics is not a skillset that academics, or even civil society, naturally have. We’ve got to educate people. And we’ve got to educate ourselves on how to effectively educate others.”

Interviewees also noted that funders could play a significant role in encouraging researchers to conduct their research ethically. “Going back to open science principles, if a group or funder is going to require or even encourage these types of data sharing practices, it’s a good idea to provide clear guidance and maybe even training as to what that looks like,” noted one of the researchers.

## Limited Infrastructure

While often not considered together, shared research infrastructure (referring to the tools and systems academics use to conduct research) and research ethics *must be coordinated*. Shared research infrastructure that is adopted by many researchers can not only facilitate more and higher quality research, but it can also reinforce disciplinary norms around what is considered robust and ethical research. For example, the academically run American National Elections Survey has become a gold standard for survey methodology,<sup>43</sup> both providing data and establishing protocols about how survey research should be done practically and ethically.

In contrast, limited platform research infrastructure makes it more difficult to build the aforementioned consensus around ethical procedures and review. This is the situation that academics who conduct platform research find themselves in.

Some of this may be a consequence of how funding is typically provided to research projects. In particular, interviewees noted that funding for research tends to be project-by-project. One researcher emphasized that “infrastructure work and long-term solutions go beyond three- or four-year projects.” However, for sustainable infrastructure that supports a field of research, longer-term funding is necessary.

For researchers, infrastructure often constitutes one of two things: technical infrastructure (e.g., data archives and programming packages) and administrative infrastructure (e.g., ethical reviews, policy, and legal frameworks). While these two types of infrastructure vary greatly, both in terms of structure and function, they often work in tandem. Both forms of infrastructure also share the same challenges of required regular maintenance.

One researcher who has helped build data infrastructure lamented that “infrastructure and maintenance are not sexy and nobody wants to pay for them.” As a result, “everyone’s just collecting these kinds of slivers of data sets that are all the same.” This results in researchers repeating efforts rather than working together. The same applies to when

these researchers then try to share data with one another, as there is as little consensus about the “right” way to provide access. Another researcher explained that at present, “Researchers collecting and sharing data can set ethical standards because they facilitate access.”

Sharing similar sentiments, a researcher who worked on software infrastructure also emphasized that building tools and packages is important for making data processing and analytical steps replicable. However, like other infrastructure, software must be regularly maintained, “If you have something that you want to run continuously, you have to fund it continuously. And that isn’t a great fit for a lot of academic projects, which are seen as having firm start and end points.”

Another important challenge to building and maintaining ethical infrastructure for platform research is the rapid development of the field, which has required routine and repeated changes to how platform research is conducted. One communication scholar described platform research as “a moving target,” making it challenging to produce software, which would require regular maintenance. A computer scientist shared similar sentiments:

“Infrastructure-focused projects may lead to more standardization, but without maintenance, these can get old quickly. Right now, everyone’s thinking about Twitter. But if we build around this, who knows what will happen? I always give the example of people who graduated with me, that their thesis was on MySpace. Who remembers that now?”

Collectively, our interviews suggest that researchers understand the many challenges to building ethical frameworks for platform research. While some researchers have, over time, determined ethical practices that work for their own research, many also expressed a desire to receive feedback, support, or recommendations for how to improve both their ethical framework and ethical practices. In lieu of a review board (or even “just someone to talk to about ethics,” as one researcher expressed), academics were instead left to make their own, highly individualized, ethical decisions.

## **BUILDING ETHICAL STRATEGIES FOR ACADEMICS**

The development of research strategies and research ethics for studying digital platforms must go hand in hand. Thus far, building an ethical standard has been mired in (1) disagreements about what research practices are considered ethical (or not), (2) a lack of consensus about procedures and ethical review, and (3) a lack of research infrastructure that provides widespread data access while also adhering to field-aligned ethical standards.

However, academic researchers express a strong desire to be ethical in their work. For platform researchers, thinking through proactive ethical strategies is more desirable than reactive ethical strategies. For this reason, many have developed ethical approaches beyond what is expected of them in their fields and institutions. This research is critical for proactively building ethical practices.

Given the variety of methods that are used to study digital platforms, there is no singular ethical framework that will work for every study. For example, researchers collecting user-consumption behavior may take a more privacy-oriented approach, whereas researchers studying political actors and celebrities may take a more public-oriented ethical framework. As described succinctly in one paper title, “We Aren’t All Going to Be on the Same Page about Ethics.”<sup>44</sup>

Based on the interviews conducted, consensus must be built around two ethical aspects: ethical review and procedural ethics. Ethical review refers to a consideration of ethics as a series of steps or a checklist. While acknowledging that this is a “minimal” consideration of ethics, such an approach is necessary to determine what is fundamentally not acceptable. As one computer scientist expressed, “We can’t ignore the value of formal guidelines and a formal process, also because we know that some communities that’s what they’re going to want.” At present, U.S. researchers broadly have relied on Institutional Review Boards as a form of ethical review, though (as noted above), they are not presently fit to evaluate platform research ethics.

The second ethical aspect is procedural ethics, referring to how researchers should pragmatically conduct ethical research. This is particularly important for researchers seeking informed consent and seeking to anonymize their content. One researcher who cares deeply about anonymization succinctly explained, “If the question is if you’re going to anonymize data... if you’re going to engage in the process of disguising information, you should do it right.”

Now is as good of a time as any to make progress on both ethical review and procedural ethics processes. Speaking optimistically about platform ethics, one researcher suggested, “If we could have a big sky project ... to really push this forward. I think it can make a big difference. And the timing is really right now because I think there’s just the right amount of public attention to [data ethics] that this could be really helpful.”

For a consensus on ethical review and procedural ethics, it is necessary for academics across multiple disciplines, using a variety of methods, to (1) conduct ethics research to determine best practices, (2) engage in coalition building to discuss how to implement these best practices pragmatically, and (3) support ethically-informed infrastructures that serve as leaders within the field.

## Conduct Ethics Research

Ethics research is a critical component of platform studies. Academics use a variety of methods to study research ethics, including both qualitative and quantitative approaches. However, the degree to which ethics research is respected in a discipline varies. Whereas a consideration of ethics may be expected in some disciplines, some researchers noted that ethics research is not as well received in their fields. Funding for this research could signal the importance of these topics. As one computer scientist noted, “I would love funding for empirical work on actual ethical practices, because that would help me incentivize my students to get involved ... But it also helps my department understand that the work that I do is valuable.”

To be clear, there *are* research ethics projects being funded. One important and ongoing project is [PERVADE](#), an acronym for “Pervasive Data Ethics”. PERVADE conducts studies with digital media users, IRBs, and computing research communities to develop best practices for research ethics. Furthermore, interviewees who identified as ethics researchers noted that they will continue to do empirical research because they believe so strongly in the importance of developing research ethics. However, funding can help advance ethics research across a larger variety of disciplines.

Additional meta-research about research ethics can also contribute to both the development of an ethical review and to procedural ethics. For example, researchers found that users on a social media platform were particularly concerned about messages that were private or had personally identifying information.<sup>45</sup> This can inform ethical reviews regarding which projects are of a higher ethical risk and which projects are of a lower ethical risk.

In terms of procedural ethics, ethical studies with academic and independent researchers can help build a process for how a researcher should review their data. For example, if a researcher wanted to anonymize or disguise their data, there should be a systematic procedure for how to do so. While some studies have developed approaches for de-identifying digital networks,<sup>46</sup> users,<sup>47</sup> and even content,<sup>48</sup> these practices have yet to be widely implemented.

This brings us to a related point: Ethics research should be better incorporated into researcher education. At present, this is done piecemeal, depending both on the field and on the extent to which ethics matters to the specific teacher of a course. However, studies of ethics pedagogy can help inform educators about how ethics can be incorporated into other classes. One researcher highlighted the importance of ethics research for STEM in particular, “I want more research focused on STEM education and ethical training ... Especially more action-oriented research; what we’re actually implementing and doing evaluations in terms of educational outcomes.”



## Coalition Building

Another important strategy is to engage in coalition building and to support convenings, both within and across disciplines. Existing professional conferences – such as the Association of Internet Researchers (AoIR), the Conference on Human Factors in Computing Systems (CHI), the International Conference on Computational Social Science (IC2S2), the International AAAI Conference on Web and Social Media (ICWSM), the Institute of Electrical and Electronics Engineers (IEEE), the International Communication Association (ICA), the National Communication Association (NCA), and the American Political Science Association (APSA) – can and have played an important role in setting ethical standards for their respective disciplines. For example, AOIR has an established ethical guideline, which has been updated three times. Some conferences, such as ICWSM and CHI, have ethical reviews.

Ethics have certainly played a critical role in several convenings. AOIR, for example, has historically hosted several panels on research ethics each year. The [Digital Data Conference](#) in 2022 consisted of both an ethics-focused day and a practice-focused day. And the [2023 post-API conference](#), which focused on data access, included a session on data ethics. However, ethics is often just one part of these events, rather than the subject that takes center stage.

In-person convenings that focus on ethical reviews are essential for discussing difficult decisions and building trust within the field of platform research. Of our 18 participants, 10 mentioned the need for an ethics-focused conference. Some researchers said that such a gathering could help turn ethical frameworks into tangible practices, “If you could get a meeting of the minds ... and set up the workshop to bring people who can write and make out pragmatic real outcomes of a list [of ethical norms].”

One researcher noted that an ethics-focused gathering would also create space for researchers from a variety of fields to come together and exchange experiences, “A conference or a workshop [as a] hub of some sort. [It would] help grow the community who are interested in these ethical questions to bring in more diversity of perspectives and solve some of these hard challenges.” This is necessary for any ethical review or procedure to be adopted by a wide range of fields. Another researcher expressed a desire to hear experiences from other researchers, particularly those studying ethics and best practices:



I think we need to spend more time with researchers who are doing [ethics] work, talking about things and sharing these stories. I learn a lot from other researchers’ experiences ... And reflecting on these challenges would be extremely productive and helpful for our community. I don’t do ethics research, but the more I hear about this work and the more space there is for people to share and reflect, it can help inform my own research processes.

In addition to physical gatherings, digital coalitions are also critical for facilitating continued conversation about ethics, regardless of where researchers are located. One researcher noted that coalition building for consensus around research ethics is important in the United States: “Organizing and field-building is not as essential, because many of the processes are centralized in [location outside of the U.S.]. Whereas in the U.S., to get any of these things done, you have to build that coalition. You’ve got to have cooperation.” To be clear, this is not to say that ethics-building outside of the United States is by any means easy; however, within the United States, researcher “buy-in” is predicated on consensus-building rather than centralized procedures.

Digital coalitions and virtual events can also help sustain conversations around ethics, particularly for those who do not have the financial resources, or time, to attend in-person convenings. “Academic convenings where we can develop a shared community around dataset uses and issues in social behavior can happen both online and offline,” one researcher from an information school expressed. They also noted that, when done virtually, such events should also include other groups, like “nonprofits and program officers, as they have so much knowledge.”

Coalition building is especially important for researchers to receive feedback. For example, one researcher indicated that they would want IRBs to be better prepared to discuss ethical challenges related to platform research, “Another thing that would make IRBs more effective is if we had consensus around what the ‘right’ sort of ethical behaviors are. It’s nice to have someone to go talk to about [ethical topics].” Other interviewees also expressed a desire to talk to a third party, such as a research ethicist, someone in IRB, or an ethics committee of a professional organization.

While some coalitions may be specific to academic researchers, or researchers alone, some participants also expressed an interest in building coalitions with, when possible or feasible, other stakeholders. Of course, there was also disagreement here: collaborative researchers were more likely to include social media representatives in convenings and coalitions, while independent researchers generally did not want these stakeholders present at events.



## Research Infrastructure as Ethical Leadership

Professional organizations and convenings play a substantive role in building consensus and encouraging adoption of ethical practices. However, research infrastructure developed for widespread use by academics can lead others to implement ethical practices. As platform research infrastructures become more commonplace, it is necessary for these infrastructures to balance open science practices with user privacy, particularly if they plan to support both privacy-oriented and public-oriented research. In their work, academics conducting platform research rely on a variety of infrastructures to do their research, including both policy infrastructure to support ethical reviews and technical infrastructure to provide responsible access to researchers.

### Ethical Technical Infrastructure

For technical infrastructure, researchers noted that the maintenance of these resources is costly but under-supported. One communication scholar argues, “Academia is in dire need of standardization ... It facilitates the evaluation of research practices, it makes research replication, it makes data reuse easier, it also reduces duplication of efforts.” For example, programming packages that are used to anonymize or standardize data can be very helpful for facilitating consistent anonymization practices across disciplines. However, such packages require regular maintenance, particularly if there are changes to the base programming language.

Researchers, and archivists in particular, also noted that data infrastructure could encourage ethical data reuse and minimize the number of novel new datasets constructed to study a social phenomenon.<sup>49</sup> As one researcher explained, “Many of the phenomena that we want to study in social media ... you don’t need today’s data, you could use data from two years ago and it wouldn’t be that different.” In this situation, archives and data access tools can play a critical role in balancing democratic access to platform data while minimizing the amount of data needed to conduct research.<sup>50</sup>

Several interviewees also noted that technical and data infrastructure can facilitate more ethical research practices by both providing data access and encouraging proactive ethical practices. While the perspective that research data should be openly accessible is admirable, there is the potential for even public platform data to be misused.<sup>51</sup> Continuing their thought processes regarding infrastructure and ethics, one researcher also noted, “If you’re going to be making data publicly available, you should go through some checklist to ensure that there is some type of vetting ... It’s great to say we’re going to require people to share data, but you need to really unpack what that means.” This suggests that data and technical infrastructure could, and should, develop ethical practices for sharing.

## Policy and Legal Infrastructure

This leads to yet another type of infrastructure: policy and ethical review. As explained earlier, our interviewees varied in whether they felt current ethical review structures (e.g., IRBs) could be modified to account for the nature of platform research. Some participants suggested an alternative ethical review procedure that was focused on platform research, such as a technology ethics board.<sup>52</sup> One computer scientist, for example, noted that one of their academic conferences conducted ethics reviews:

“I’m on the [conference] ethics board and, if a paper is flagged for an ethics review, we conduct an external review. We basically come in as a fourth reviewer, only commenting on the ethics of the research, what we think should be done, and whether or not they breached normative ethical guidelines within our community. [...] But [this conference] is the only one I know that has a full-on ethics committee.”

Others suggested modifications or addendums to the Belmont Report that could adjust the report’s principles (respect for persons, beneficence, and justice) in the context of platform research. Focusing primarily on social benefits, one researcher said, “The Belmont Report provides some sort of guidance for how we’re supposed to think about [research], but it doesn’t say anything about communities or social benefits.” They continued, “Individual harms are important, but when we look at this through an individual control lens, we might ignore community membership harms and risks, and I don’t know if we have a good way for thinking about harms to communities in a way that is analogous to harms to the individual.”

Going beyond this, two researchers noted that legal policies and frameworks around platform research ethics and data access would facilitate more responsible, ethical, and standardized research practices. Pursuing policy change proactively, rather than reactively (i.e., after an ethical crisis) was particularly important to several interviewees: “The legal part is a necessary criterion, especially if there are things that should absolutely not be done.”

The risks of not having an ethical or legal framework should not be understated, for both individual academics and their institutions. Rather than having a procedure or process in place, researchers are left to develop their own practices piecemeal. As one communication researcher explained, “We’ve had a lot of experiences doing contract negotiations with [companies] and it becomes a lot of legal work for us.” This creates inconsistencies across institutions and disciplines, producing different levels of risk for different researchers, and potentially chilling important research. Without some set of guidelines about what is or is not acceptable, “my work is more at the whims of what a platform thinks is okay.”

Here, too, funding can play a critical role. One interviewee, who has relied traditionally on federal grants, lamented, “I’ve never seen a federal grant that has included support for legal issues.” Importantly, funders have already begun to play a critical role in this space by providing resources for legal counsel, particularly as researchers become targets for harassment for conducting their work.

While these recommendations seem lofty, and they are, researchers also expressed optimism that support for ethical research, coalition building, and research infrastructure can assist platform research. As one researcher enthusiastically said, “Perhaps my silver lining or my optimistic view is that there’s certainly a demand from researchers for [an ethical framework], even if there isn’t the institutional framework quite yet.”

## **CONCLUSION AND RECOMMENDATIONS**

At present, U.S. digital platform researchers are caught between two principles: producing transparent, robust research that benefits the public and protecting a user’s right to privacy. While an emphasis on one over the other may be clear in some cases, there are many circumstances that require a more nuanced ethical practice that combines both public and privacy considerations.

As identified by the interviews and a review of the literature, it is exactly these nuances that researchers are struggling with. The challenge of balancing these principles is exacerbated by both abundant and sometimes conflicting guidance from platforms and legal counsel, as well as by a lack of a proper ethical review. This is not sustainable, particularly given that different research methods, studying different contexts, will likely apply different ethical frameworks.

Interviewees highlighted the need for more consistent ethical review, coalition building, and support for technical, data, and policy infrastructure. These findings lead to several recommendations:

### **Ethics Research**

- More funding and support for ethics research is essential, particularly for conducting ethical reviews and developing ethics procedures.
- Professional academic organizations should recognize and reward ethics research in their field.
- Professional academic organizations should adopt ethical guidelines.

## **Coalition Building**

- There should be more interdisciplinary coalition building.
- Research associations should develop ethical guidelines for researchers within their discipline.
- Researchers should organize in-person convenings that are dedicated to platform research ethics.

## **Research Infrastructure**

- There should be a platform research or digital data ethical review that operates within or in tandem with Institutional Review Boards.
- Funding for research infrastructure should ask about the collaborative or independent nature of the research, as well as whether the infrastructure uses a privacy-oriented, public-oriented, or hybrid approach to research ethics.
- There should be funding for long-term infrastructure, including, but not limited to, tools for anonymization, data archives, and benchmark datasets.
- Researchers should engage with policymakers to develop a legal framework for platform research that can govern data access and set boundaries regarding what researchers can do with the data.

As studying digital platforms and social media becomes increasingly popular, researchers, platform users, companies, politicians, and funders must work together to support ethical research practices that are flexible to the ever-changing field, and yet guided by the shared principles of research for the public and minimizing user harm.

## METHODOLOGY

The 17 interviews with 18 academic researchers took place over four months. Each interview lasted approximately one hour. While there is growth in the field, it remains small enough that demographic information may be personally identifying. However, we note that of the 18 interviewees, eight identified as women or non-binary, and three identified as people of color. Only one interviewee was currently working outside of the United States, but they had conducted research in the United States previously (two other interviewees had joint appointments in non-U.S. institutions). All the interviewees study digital platforms, most frequently with quantitative or computational methods; although several researchers used qualitative methods such as digital ethnographies.

### Interviewee Information

	Identifier	Field (Department)	Rank
1	A	Communication	Full
2	B	Communication	Associate
3	C	Computer Science	Assistant
4	D	Computer Science	Assistant
5	E	Psychology	Associate
6	F	Communication	Associate
7	G	Political Science	Full
8	H	Computer Science	Full
9	I	Economics	Associate
10	J	Communication	Associate
11	K	Information School	Associate
12	L	Computer Engineering	Full
13	M	Computer Science	Staff
14	N	Information School	Associate
15	O	Information School	Associate
16	P	Political Science	Associate
17	Q	Psychology	Associate
18	R	Communication	Associate

These interviews were combined with a review of the literature on research ethics across multiple fields. While literature on data ethics in medical studies and social sciences was reviewed, the focus is more on recent literature (within the past seven years) published in the social sciences, digital humanities, and applied STEM fields. Also reviewed were professional ethical guidelines – including the [Association of Internet Researcher’s \(AoIR\) Internet Research Ethics 3.0](#), the [American Political Science Association’s Principles and Guidance](#) document, and the [ACM Code of Ethics and Professional Conduct](#) – as well as research reports such as the [Center for Democracy and Technology’s Defending Data](#) report and the [European Digital Media Observatory’s](#) report on platform-to-researcher data access.

Drawing from these readings and the interviews, open coding was then conducted to identify the ethical challenges and solutions that academics highlighted in their interviews or in their writing. This approach is in line with other qualitative analyses.<sup>53</sup> Drawing from these codes, three broad themes discussed in this report were identified: (1) ethical practices in current research designs, (2) challenges to building ethical frameworks, and (3) building ethical strategies. Additional citations and references can be found in this [Resource List](#).

## ACKNOWLEDGMENTS

First and foremost, this work could not be done without the interview participants, who generously volunteered their time and shared their experiences. Platform research is better because of them. I am also grateful to Bin Chen, who helped conduct some interviews. I am also grateful to the Media Democracy Fund for coordinating the project, and the Netgain Partnership for their continued support. Finally, I would like to thank MDDC members Megan Brown (University of Michigan), Kaiya Soorholtz (Center for Media Engagement), Jiyoun Suk (University of Connecticut), Yini Zhang (Buffalo University), and Meredith Pruden (Kennesaw State University) for their feedback on earlier drafts. All research, findings, and recommendations presented here represent the independent analysis of the author and are not intended to communicate the views of the NetGain Partnership, which commissioned this report.

## ENDNOTES

- <sup>1</sup> For more, see Shilton, K., & Sayles, S. (2016, January). “We Aren’t All Going to Be on the Same Page about Ethics”: Ethical Practices and Challenges in Research on Digital and Social Media. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1909-1918). IEEE. <https://ieeexplore.ieee.org/abstract/document/7427422>
- <sup>2</sup> <https://www.netgainpartnership.org/resources/2021/2/25/new-approaches-to-platform-data-research>
- <sup>3</sup> Shapiro, E. H., Sugarman, M., Bermejo, F., & Zuckerman, E. (December 8, 2021). New approaches to platform data research. <https://drive.google.com/file/d/1bPsMbaBXAROUYVesaN3dCtfaZpXZgl0x/view>
- <sup>4</sup> Veglis, A. (2014). Moderation techniques for social media content. In *Social Computing and Social Media: 6th International Conference, SCSM 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings 6* (pp. 137-148). Springer International Publishing. [https://link.springer.com/chapter/10.1007/978-3-319-07632-4\\_13](https://link.springer.com/chapter/10.1007/978-3-319-07632-4_13)
- <sup>5</sup> Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford University Press.
- <sup>6</sup> Berryman, C., Ferguson, C. J., & Negy, C. (2018). Social media use and mental health among young adults. *Psychiatric quarterly*, 89, 307-314. <https://link.springer.com/article/10.1007/s11126-017-9535-6>
- <sup>7</sup> <http://research-ethics.org/>
- <sup>8</sup> Israel, M., & Hay, I. (2006). *Research ethics for social scientists*. Sage.
- <sup>9</sup> Baum, M. (1994). Informed consent. Reactionary approach inhibits progress. *BMJ: British Medical Journal*, 308(6923), 271. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2539296/>
- <sup>10</sup> Metcalf, Jacob, and Kate Crawford. “Where are human subjects in big data research? The emerging ethics divide.” *Big Data & Society* 3, no. 1 (2016): 2053951716650211. <https://journals.sagepub.com/doi/full/10.1177/2053951716650211>
- <sup>11</sup> Kia-Keating, Maryam, Diana Santacrose, and Sabrina Liu. “Photography and social media use in community-based participatory research with youth: Ethical considerations.” *American journal of community psychology* 60, no. 3-4 (2017): 375-384. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajcp.12189>; Matias, J. Nathan, and Merry Mou. “CivilServant: Community-led experiments in platform governance.” In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1-13. 2018. <https://dl.acm.org/doi/abs/10.1145/3173574.3173583>
- <sup>12</sup> Doerfler, P., Forte, A., De Cristofaro, E., Stringhini, G., Blackburn, J., & McCoy, D. (2021). “I’m a Professor, which isn’t usually a dangerous job”: Internet-facilitated Harassment and Its Impact on Researchers. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-32. <https://dl.acm.org/doi/abs/10.1145/347608>
- <sup>13</sup> Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4), 665-668. <https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1477506>
- <sup>14</sup> <https://www.texastribune.org/2023/11/15/texas-tiktok-ban-university-hearing/>
- <sup>15</sup> Lukito, J., Brown, M. A., Dahlke, R., Suk, J., Yang, Y., Zhang, Y., ... & Soorholtz, K. (2023). The State of Digital Media Data Research, 2023. *UT Faculty/Researcher Works*. <https://mddatacoop.org/dmd/>; Jimenez-Marquez, J. L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J. L., & Ruiz-Mezcua, B. (2019). Towards a big data framework for analyzing social media content. *International Journal of Information Management*, 44, 1-12. <https://www.sciencedirect.com/science/article/abs/pii/S0268401218305073>
- <sup>16</sup> D’Arcy, A., & Young, T. M. (2012). Ethics and social media: Implications for sociolinguistics in the networked public 1. *Journal of Sociolinguistics*, 16(4), 532-546. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9841.2012.00543.x>; Kessler, M., Marino, F., & Liska, D. (2023). Netnographic research ethics in applied linguistics: A systematic review of data collection and reporting practices. *Research Methods in Applied Linguistics*, 2(3), 100082. <https://www.sciencedirect.com/science/article/pii/S2772766123000423>



- <sup>17</sup> Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, *11*(10), e1004513. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004513>; Andreotta, M., Nugroho, R., Hurlstone, M. J., Boschetti, F., Farrell, S., Walker, I., & Paris, C. (2019). Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. *Behavior research methods*, *51*, 1766-1781. <https://link.springer.com/article/10.3758/s13428-019-01202-8>
- <sup>18</sup> Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, *595*(7866), 189-196. <https://www.nature.com/articles/s41586-021-03660-7>
- <sup>19</sup> King, K. (2011). Professional learning in unlikely spaces: Social media and virtual communities as professional development. *International Journal of Emerging Technologies in Learning (IJET)*, *6*(4), 40-46. <https://www.learntechlib.org/p/45112/>; Lefebvre, H., Legner, C., & Fadler, M. (2021, December). Data democratization: toward a deeper understanding. In *Proceedings of the International Conference on Information Systems (ICIS)*. [https://aisel.aisnet.org/icis2021/gen\\_topics/gen\\_topics/7/](https://aisel.aisnet.org/icis2021/gen_topics/gen_topics/7/)
- <sup>20</sup> Samuel, G., & Buchanan, E. (2020). Guest editorial: Ethical issues in social media research. *Journal of Empirical Research on Human Research Ethics*, *15*(1-2), 3-11. <https://journals.sagepub.com/doi/full/10.1177/1556264619901215>
- <sup>21</sup> Markham, A. N., Tiidenberg, K., & Herman, A. (2018). Ethics as methods: doing ethics in the era of big data research—introduction. *Social Media+ Society*, *4*(3), 2056305118784502. <https://journals.sagepub.com/doi/full/10.1177/2056305118784502>
- <sup>22</sup> <https://research.facebook.com/2020-election-research/>
- <sup>23</sup> Wagner, M. W. (2023). Independence by permission. *Science*, *381*, 388-391. <https://www.science.org/doi/10.1126/science.adi2430>
- <sup>24</sup> Brown, M. (2023, March 1). “The Problem with TikTok’s New Researcher API is Not TikTok” *Tech Policy Press*. Retrieved from <https://www.techpolicy.press/the-problem-with-tiktoks-new-researcher-api-is-not-tiktok/>
- <sup>25</sup> Bak-Coleman, J. (2023, February 22). “TikTok’s API Guidelines Are a Minefield for Researchers” *Tech Policy Press*. Retrieved from <https://www.techpolicy.press/tiktoks-api-guidelines-are-a-minefield-for-researchers/>
- <sup>26</sup> Lukito, J. Mathias, N., & Gilbert, S. (2023, May 10). Enabling Independent Research Without Unleashing Ethics Disasters. *Tech Policy Press*. Retrieved from <https://www.techpolicy.press/enabling-independent-research-without-unleashing-ethics-disasters/>
- <sup>27</sup> van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Communication Methods and Measures*, *16*(4), 266-282. <https://www.tandfonline.com/doi/full/10.1080/19312458.2022.2109608>
- <sup>28</sup> Marres, N., & Weltevrede, E. (2013). Scraping the social? Issues in live social research. *Journal of cultural economy*, *6*(3), 313-335. <https://www.tandfonline.com/doi/full/10.1080/17530350.2013.772070>
- <sup>29</sup> For example, in July 2023, Twitter/X owner [Elon Musk sued four unidentified individuals](#) for scraping platform data.
- <sup>30</sup> Davidson, B. I., Wischerath, D., Racek, D., Parry, D. A., Godwin, E., Hinds, J., ... & Cork, A. G. (2023). Platform-controlled social media APIs threaten Open Science. *Nature Human Behaviour*, *7*(12), 2054-2057. <https://www.nature.com/articles/s41562-023-01750-2>
- <sup>31</sup> Fiesler, C., Beard, N., & Keegan, B. C. (2020, May). No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In *Proceedings of the international AAAI conference on web and social media* (Vol. 14, pp. 187-196). <https://ojs.aaai.org/index.php/icwsm/article/view/7290>
- <sup>32</sup> Chua, S. M. (2022). Navigating conflict between research ethics and online platform terms and conditions: a reflective account. *Research Ethics*, *18*(1), 39-50. <https://journals.sagepub.com/doi/full/10.1177/17470161211045526>



- <sup>33</sup> Vitak, J., Shilton, K., & Ashktorab, Z. (2016, February). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (pp. 941-953). <https://dl.acm.org/doi/pdf/10.1145/2818048.2820078>
- <sup>34</sup> Roozenbeek, J., & Zollo, F. (2022). Democratize social-media research-with access and funding. <https://www.nature.com/articles/d41586-022-04407-8>
- <sup>35</sup> Shilton, Katie, and Sheridan Sayles. "We Aren't All Going to Be on the Same Page about Ethics": Ethical Practices and Challenges in Research on Digital and Social Media." In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 1911. IEEE, 2016. <https://ieeexplore.ieee.org/abstract/document/7427422>
- <sup>36</sup> Zimmer, M. (2020). "But the data is already public": on the ethics of research in Facebook. In *The ethics of information technologies* (pp. 229-241). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003075011-17>
- <sup>37</sup> Shilton & Sayes, 2016.
- <sup>38</sup> Do, K., Pang, R. Y., Jiang, J., & Reinecke, K. (2023, April). "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). <https://dl.acm.org/doi/abs/10.1145/3544548.3581347>
- <sup>39</sup> Gangarde, R., Shrivastava, D., Sharma, A., Tandon, T., Pawar, A., & Garg, R. (2022). Data anonymization to balance privacy and utility of online social media network data. *Journal of Discrete Mathematical Sciences and Cryptography*, 25(3), 829-838. <https://www.tandfonline.com/doi/pdf/10.1080/09720529.2021.2016225> ; Weinhardt, M. (2021). Big data: Some ethical concerns for the social sciences. *Social Sciences*, 10(2), 36. <https://www.mdpi.com/2076-0760/10/2/36>
- <sup>40</sup> Tsan, M. F. (2019). Measuring the quality and performance of institutional review boards. *Journal of Empirical Research on Human Research Ethics*, 14(3), 187-189. <https://journals.sagepub.com/doi/abs/10.1177/1556264618804686>
- <sup>41</sup> Zimmer, 2017.
- <sup>42</sup> Raymond, N. (2019). Safeguards for human studies can't cope with big data. *Nature*, 568(7753), 277-278. <https://www.nature.com/articles/d41586-019-01164-z>
- <sup>43</sup> Krosnick, J. A., & Lupia, A. (2012). The American National Election Studies and the Importance of New Ideas. *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*, 9.
- <sup>44</sup> Shilton, K., & Sayles, S. (2016, January). "We Aren't All Going to Be on the Same Page about Ethics": Ethical Practices and Challenges in Research on Digital and Social Media. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 1909-1918). IEEE. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7427422>
- <sup>45</sup> Fiesler & Proferes, 2018; Hemphill et al., 2022
- <sup>46</sup> Fu, L., Zhang, J., Wang, S., Wu, X., Wang, X., & Chen, G. (2020). De-anonymizing social networks with overlapping community structure. *IEEE/ACM Transactions on Networking*, 28(1), 360-375.. <https://ieeexplore.ieee.org/abstract/document/8967164>
- <sup>47</sup> Boeschoten, Laura, Roos Voorvaart, Ruben Van Den Goorbergh, Casper Kaandorp, and Martine De Vos. "Automatic de-identification of data download packages." *Data Science* 4, no. 2 (2021): 101-120. <https://content.iospress.com/articles/data-science/ds210035>
- <sup>48</sup> Reagle, J. (2022). Disguising Reddit sources and the efficacy of ethical research. *Ethics and Information Technology*, 24(3), 41. <https://link.springer.com/article/10.1007/s10676-022-09663-w>
- <sup>49</sup> A similar point is made in [Vogus, 2023](#).
- <sup>50</sup> Roozenbeek, Jon, and Fabiana Zollo. "Democratize social-media research-with access and funding." (2022): 404-404. Retrieved from: <https://iris.unive.it/handle/10278/5010400>

<sup>51</sup> See [Zimmer, 2018](#), for an example.

<sup>52</sup> Pagoto, S., & Nebeker, C. (2019). How scientists can take the lead in establishing ethical practices for social media research. *Journal of the American medical informatics association*, 26(4), 311-313. <https://academic.oup.com/jamia/article/26/4/311/5304024>

<sup>53</sup> Marland, A., & Esselment, A. L. (2019). Negotiating with gatekeepers to get interviews with politicians: Qualitative research recruitment in a digital media environment. *Qualitative Research*, 19(6), 685-702. <https://journals.sagepub.com/doi/abs/10.1177/1468794118803022>