The Computing Research Association (CRA)'s Computing Community Consortium (CCC) and CRA-Industry (CRA-I) Response to the National Telecommunications and Information Administration (NTIA), Department of Commerce's Request for Comments: Ethical Guidelines for Research Using Pervasive Data

January 15, 2025

Written by: Nazanin Andalibi (University of Michigan), David Danks (University of California, San Diego), Haley Griffin (Computing Research Association), Mary Lou Maher (Computing Research Association), Jessica McClearn (Google), Chinasa T. Okolo (The Brookings Institution), Manish Parashar (University of Utah), Jessica Pater (Parkview Health), Katie Siek (Indiana University), Tammy Toscos (Parkview Health), Helen V. Wright (Computing Research Association), and Pamela Wisniewski (Vanderbilt University)

This response is from Computing Research Association (CRA)'s Computing Community Consortium (CCC) and CRA-Industry (CRA-I). CRA is an association of nearly 250 North American computing research organizations, both academic and industrial, and partners from six professional computing societies. The mission of the CCC, a subcommittee of CRA, is to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges. The mission of CRA-I, another subcommittee of CRA, is to convene industry partners on computing research topics of mutual interest and connect them with CRA's academic and government constituents for mutual benefit and improved societal outcomes.

Please note any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the authors' affiliations.

Below we respond to questions 1-9, 11, and 13-14 of the Request for Comments.

1. What are the potential benefits of developing national-level ethical guidelines for researchers collecting, analyzing, and sharing pervasive data?

National-level guidelines for researchers that use pervasive data would help facilitate accountability and a standard for research that could support both academic and industry

researchers, and enhance public trust in how their data are used. These guidelines could ensure a consistent approach to research ethics across different institutions, including but not limited to, universities, technology companies, and industries (e.g. healthcare, education, transportation). Researchers who work with sensitive, pervasive data would be able to use these guidelines to improve and justify their methods. This would prevent individual institutions from having varied interpretations that could either prevent important research or encourage unethical practices. This guidance would be important for all researchers, but especially those who have less support, training, or resources at their local institutions.

The guidelines also have the opportunity to protect researchers who study controversial topics by providing ethical foundations for their work. National guidelines could ensure that pervasive data involving human subjects undergoes proper ethical review. It would also ideally protect the participants whose data is being analyzed and strengthen the protections under Common Rule because many Institutional review boards (IRBs) deem some pervasive data studies as "non-human subject research," which does not ensure the same level of ethical review as research deemed human subjects research. Developing these guidelines would ensure that studies that do not go through IRB review could still be evaluated for ethical implications.

Companies can also improve their practices using these guidelines. There is currently a substantial lack of clarity about potential legal exposure for companies due to re-identifiability and other potential privacy infringements. While these guidelines would obviously not be legally binding, they can help to clarify some of the key concepts and practices. More generally, universally recognized guidelines could help to standardize their research processes, codify best practices, make their work more equitable (by ensuring all stakeholders are considered), and likely improve the efficacy of their research as well.

2. What are the potential drawbacks of developing national-level ethical guidelines for researchers collecting, analyzing, and sharing pervasive data?

The potential drawbacks include: 1) guidelines would need to stay robust to shifts in data practices; 2) there is no enforcement mechanism to ensure good faith adoption among researchers; 3) guidelines may restrict flexibility and innovation of ethical research that falls outside the existing guidelines; 4) data and researchers can both be outside of the U.S., and national guidelines need to consider international contexts.

Digital technologies are constantly evolving. Our understanding of their ethical implications and need for governance are regularly fluctuating, and new laws and bills are proposed at local, national, and international levels regularly. There would need to be a regular (perhaps annual or

biannual) review of the guidelines to ensure alignment with current policies and needs. The guidelines should be robust to specific technologies and developments, so that they can be relevant, useful, and impactful, even as the data collection and analysis technologies change over time.

Another concern is that non-binding guidelines are typically used and followed only by more-diligent researchers. As a result, the communities and people that should be protected and treated ethically may still face significant risks.

In addition, researchers using pervasive data may not be aware of the guidelines. Awareness could be increased through integration with IRB CITI or NSF RCR training and by obtaining buy-in from organizations that publish research (e.g., ACM, IEEE, National Academies) or fund research (e.g., NSF, NIH, The Knight Foundation). For industry researchers, companies would need to adopt and integrate the guidelines into their own internal practices and evaluation metrics. In some cases, researchers may be aware of guidelines but misuse them to justify unethical data practices. They might also design research studies to fall outside the boundaries of such guidelines (similar to researchers trying to avoid IRB review). The proposed guidelines should be developed in parallel with other academic and industry standards (e.g., IRB, IEEE) to minimize misuse and establish accountability when it occurs. The NTIA should also regularly examine for consistencies and conflicts with other recognized guidelines (e.g., how does IRB treat “public data” if there are new guidelines that are misaligned with IRB).

Guidelines are designed to address the common or anticipated research practices. In some cases, the guidelines would be misaligned with certain methods, populations, or domains, where data should be treated differently than what the guidelines suggest. There should be a process for discussion and appeals, especially if the guidelines manifest into requirements. It would also be beneficial for researchers to publish their protocols of these edge cases for the guideline oversight committee to review periodically to update the guidelines.

Similarly, data is often generated, stored, or collected internationally, and data analysis practices should consider local customs, laws, and values when conducting their research. Researchers themselves in both industry and academia may have international affiliations or be conducting research internationally, and the expectations for their research practices may need to consider both local and international contexts. Guidelines should encourage consideration of international contexts, while also recognizing that potential grey areas can be used to sidestep ethical practice.

3. To what extent does the definition of pervasive data in this Request for Comments capture the appropriate scope for national ethical guidelines?

It is difficult to capture a definition for pervasive data that is simultaneously comprehensive and useful. There are concerns that the definition may not fully capture all relevant scenarios, particularly those involving data collected through networked or digital services that are not necessarily online (i.e. Internet-connected). Therefore, we encourage NTIA to use "digital services" or "networked services" rather than "online" since "online" can be interpreted as "on the internet", which is too narrow in scope. As noted above, the definition here may be difficult to interpret in tandem with the IRB, so researchers may need guidance on how to approach pervasive data about people that do not qualify as human subjects per the IRB. The emphasis on 'data about people' is important and can heighten ethical standards rather than seeing data as about `users', which can dehumanize people and may or may not be actual humans (e.g. corporate accounts; AI).

a. Are there particular types of data or other digital artifacts that should be carefully considered or included/excluded in the definition?

We suggest considering the explicit inclusion of health data, biometric data, sensor data (e.g., tracking body movements, sensed behavior of humans), non-publicly available data, personally identifiable information (PII), data of marginalized or at-risk communities (i.e., people at risk for poor health and social well being (Wisniewski et al.)), inferred data (i.e., algorithmic inferences of one's identity, activities, emotion or affect, likeness, etc.). Based on the broad definition of pervasive data, it may be beneficial to create a taxonomy of data types, akin to the General Data Protection Regulation (GDPR) 4-tier data classification, that considers privacy, benefits versus risks, sensitivity, and other considerations.

b. Are there pre-existing similar definitions, similar to the one provided, that should be considered?

The publication, *Health Data Sharing to Support Better Health Outcomes: Building a Foundation of Stakeholder Trust* (National Academy of Medicine 2020), defines health data as all the information that accumulates about a person or population that may affect health outcomes. This includes but not limited to: 1) data in an electronic health record or other storage that pertains to clinical care; 2) data gathered by health services or clinical researchers; 3) genomic data; 4) social needs data collected by government agencies or clinicians; 5) health insurance claims; 6) patient reported health outcomes data (e.g., activity tracking data), which have been gathered from individuals (patients, family members, or caregivers). Health data can be protected by removing 7 identifiers,

according to HIPAA regulations, including: patients name, social security number, medical record number, phone number, dates of care, address, biometric identifiers. We encourage guideline developers to be mindful of pervasive data related to health and the implications of that data.

4. What are some existing barriers to accessing pervasive data?

Pervasive data collected and stored in technology companies is generally inaccessible to researchers outside of those companies. The predominant challenge is the misalignment between companies' priorities (profit, legal liability) and researchers' priorities (creating new knowledge). In cases where data is available, it may be prohibitively expensive which disproportionately impacts resource-constrained researchers. Where researchers do gain access to data, there are few mechanisms to assess the quality of that data (e.g. is it accurate; what is missing; who is missing).

Pervasive data that is inaccurate and/or unverified could lead to research findings and policy implications that are misinformed. The idea that pervasive data is a robust proxy of what happens in the real world is generally problematic - as not everyone is equally represented in the digital world (e.g., consider digital divides). Specific to public-facing data (e.g., social media), it is unclear who can provide consent/permission to use the data. If researchers are to have access to pervasive data, we would need a large-scale shift in thinking about how that data is made available, what protections are available to companies, what standards researchers should be held to, and how to evaluate the quality of the data.

5. What data held by online services would be most valuable to the public interest if researchers were able to access it?

First, we must understand what the public would consider data collection and associated research that is in their interest, rather than assuming what those might be, which may vary based on identity factors, local policies, and the context of data collection and use. That said, we offer three broad data types that are important for the public interest. The first is data that aligns with societal priorities (e.g., democracy, healthcare, housing, children, poverty). This could allow researchers and companies to develop technologies and policies towards the greater good - e.g., helping those most in need. The second is data that helps us understand how technology is shaping our social lives (e.g., social media). This is important for understanding and tackling societal challenges, like disinformation, harassment, and mental health. The third is data that is collected, analyzed, and used (by various actors) about people,

without their informed consent or even awareness. This is important for protecting people's privacy and enabling them to make informed decisions about their digital behaviors, identities, and likeness.

There should be transparency about what data exists and how it can be used so that people understand it is not necessary to explicitly reveal information to give a lot of information. It would be very important in these cases that if such sensitive information is collected that it could not be used in harmful ways (e.g., surveillance, legal recourse). The public should also know what types of content is being presented/pushed to them via their timelines/ads/ accounts for their benefit and potential harm. For example, an individual in recovery from substance use disorder may receive ads or messaging that are triggering old behavioral patterns and are an impediment to recovery. If individuals were aware of how this content is being pushed to them (e.g., data provenance) it would be easier for them to engage with technology in a healthier manner. The aggregate of this information would also be helpful for researchers who are interested in understanding the relationship between engaging online content and behavior and how to adapt content based on current behavioral changes.

In national priority areas, such as youth online safety and digital well-being, data on risk-related behaviors on social media, like what processes are in place for moderation, reporting, etc., would be very useful. Therefore privacy protections and strong guidelines for how such data can and cannot be used are critical. It would also be useful to know what types of content are social platforms defining as "harmful."

Additionally, it would be helpful to have an established process for obtaining informed consent from users to collect and analyze the data through the platforms that clearly help researchers identify who has and has not consented. Consent should be truly informed (Davies, 2022) - not like a terms of service agreement that most people do not read, but need to accept to use a product. Conducting the consent process outside of the platform and requiring users to donate their data directly is costly and takes a lot of resources that could be better managed through collaboration between researchers and social media companies. This process could also increase the ethics around collecting such data.

For healthcare research, aggregate data that can be used as a proxy for social needs (e.g., percentage of reduced or free lunches in a school district, household income, etc.) would be extremely valuable when paired with data from sensing technologies (e.g., continuous glucose monitors, implanted cardiac devices, and other health monitoring technologies) for a particular area where health needs must be defined to deliver better services. This data can provide insights into public health trends and stands to improve health outcomes for individuals.

6. Consent and autonomy are key principles in human subjects research ethics. However, users of online services may be required to divulge certain personal information and/or have no ability to freely make decisions about its use. How should researchers working with pervasive data consider consent and autonomy?

People are expected to know how to keep their data safe and private online, however most people are not aware of how to do this nor can they keep up with each policy change and law (Wei et al., 2023). Indeed, if people read every privacy policy they encountered, it would take more than 244 hours per year (McDonald & Cranor)! Researchers should move away from a blanket consent model (e.g., having to consent to access a mobile app or Terms of Service). Uniform consent is not possible since behavior and expectations change across time and contexts. There should be dynamic, in-context, user-friendly privacy considerations so that the user consents to a more specific request that is reasonable for their immediate use case. Further, new paradigms and frameworks for large-scale data donation directly from users of a platform (Razi et al., 2022), rather than data scrapes of publicly or semi-publicly (i.e., accessible with the creation of an account) available data could be implemented, so that participants must opt-in, rather than be given little or no choice to opt-out.

In contexts that are imbued with power imbalances (e.g., workplace, healthcare), consent and autonomy in relation to data collection are especially challenging. While consent has always been challenging in these settings, the scale and types of harm associated with pervasive data use merits special attention. For example, would patients who are ill (e.g., mental health, chronic illness, older adult autonomy assessments) consent to digital surveillance monitoring using wearables, ambient sensing, emotion recognition systems if they truly understood how all of the "non-intrusive" data captured and the inferences drawn from that data could lead to possible harmful decisions (e.g., institutionalization, insurance increases, lack of autonomy)?

When soliciting consent, there should be clarity about if allowing access to their data is (1) legally required (e.g. age before purchasing alcohol to be delivered) or is motivated by the company's desire for data, (2) required in order to use the service/obtain the information the user is attempting to access, and (3) going to potentially be sold/accessed to a data broker and/or researchers and/or other actors (e.g., government, law enforcement).

An example of how autonomy can be honored when using pervasive data from social media sites is to use, where possible (1) publicly accessible data, and (2) data that is still live online at the time the research is being conducted. For example, guidelines should consider if researchers analyzing large-scale Reddit datasets, which may be collected prior to research, should confirm if a post has been deleted and how it would be analyzed in future publications.

Guidelines should address how to deal with data if/when people choose not to consent to data collection (e.g., a participant agreed during initial data collection, but then withdrew consent). How can this data “be forgotten” when used in algorithmic systems, qualitative analysis? People who choose not to consent or withdraw their consent should have stricter guidelines protecting their data. Even once researchers have the ability to collect and analyze user data, there needs to be protections against that data being used against the individuals.

A celebrated principle is to design for increased trust, but it is important for people to make informed assessments about what is and is not trustworthy. It is crucial to invent new ways that individuals can understand the privacy of their data (e.g., an interactive infographic or occasional surveys to understand people’s perspectives on their data being included in research).

Guidelines should provide researchers with suggestions on how to keep data safe. For example, if researchers are working with health-related data, NIH Certificates of Confidentiality are legal protections provided by the National Institutes of Health to safeguard sensitive research information from forced disclosure (e.g., via Freedom of Information Act requests). They are issued to researchers conducting studies involving identifiable, sensitive information, such as genetic data, behavioral health, or other personal details. These certificates prevent researchers from being compelled to disclose participants' data in legal proceedings, thereby encouraging honest participation and protecting participants' privacy.

The principle of “do no harm” is an alternative model to traditional consent that can provide protection for data subjects in cases where autonomy is limited or consent is given in circumstances where it is required for the individual to have access to required or desired resources. This principle can be implemented by researchers and designers of technology by acknowledging a statement similar to a code of conduct: researchers and designers should agree that they will preserve the privacy and rights of the individuals associated with the pervasive data being used for their research or in the use or application of their technology. This shifts the burden of protection on the researcher or technology designer rather than on the consent from individuals whose data the research or technology relies on.

7. What ethical issues and risks to privacy and other rights, and mitigation strategies, should be considered during the research design phase?

One of the incentives for collecting and analyzing pervasive data is that it tends to allow us to do faster research at larger economies of scale, which can be more efficient than other types of

research (e.g., interviewing people). However, there are significant ethical concerns related to doing more research faster. One of them is that given the variety of practices around the collection of pervasive data, researchers who use pervasive data should take precautions and not assume that data was collected ethically. If there are any doubts about how the data was obtained, or if the use case is appropriate, researchers should refrain from moving forward.

8. What are the risks and mitigation measures related to pervasive data acquisition and access?

a. What are the risks to data subjects resulting from the methods used by researchers to access pervasive data? How do these risks vary based on the methods of access?

Inferences made algorithmically about personally sensitive information based on pervasive data (e.g., inferences about people's health status or emotional states based on social media and/or wearable data) pose significant risks to data subjects.

c. What are the current best practices for de-identifying, pseudonymizing, or aggregating pervasive data? What practices exist to prevent or reduce the chance of re-identification of de-identified data? Where do these techniques fall short? What research questions may require identifiable data, and why?

Ethical fabrication (Markham, 2012), the process of translating or synthesizing qualitative data so that the sentiment is shared, but participants' identity is hidden (e.g., one cannot search Reddit for a quote from a paper), is commonly used in qualitative data and social analysis. Two major challenges exist with ethical fabrication (1) ensuring the original person(s)' data is accurately translated, and (2) the decreased ability to replicate/verify researchers' findings.

Some researchers do not list in their publications where they collected data, however, this makes it difficult for the broader research community to learn best practices for methods and analyze the results.

e. Under what conditions should data subjects be notified that their data is used for research? What are necessary and/or best practices for communicating with data subjects when their data is used for research? What barriers exist to notifying data subjects?

It would be good to ask the public about their expectations regarding when they would wish to be notified that their data is used for research. That said, at the very least when data collection and analysis shapes people's experiences in any way, they should be

notified. Major challenges with sharing findings with people include (1) if a research team shares their findings within a community, then others can identify the community and thus put participants at more risk for de-identification, and (2) if people were unaware that their data was used, then find out, there must be processes for their data to be withdrawn - which can impact the dissemination of results.

i. When should informed consent be obtained from users or data subjects? What should be the differences between informed consent obtained for a specific project versus for commercial or general secondary use (e.g., "broad consent")? What are the barriers to obtaining informed consent from users and data subjects?

We acknowledge that road consent practices do not acknowledge privacy as contextual and identity-dependent. The barriers for obtaining informed consent from users and data subjects include *whom* to consent *when* under *what context* based on what *potential harms.* Do volunteer forum moderators speak for all forum members? Does a company speak for all users of a given system/app?

9. What are the risks and mitigation measures that arise when processing and analyzing pervasive data?

a. Researchers will sometimes combine pervasive data with other pervasive data or with non-pervasive data from other sources. How might this impact risks? What best practices exist to mitigate these risks?

Here are a few mitigation measures to consider: (1) if a public post has been deleted, do not include it in the research, (2) review a dataset before embarking on research to ensure the autonomy of the person, (3) document the provenance of data and possible harms of using the data, (4) consider developing and using donated data repositories.

10. What are the risks to privacy and other rights related to the dissemination and archiving of research outputs? What mitigation measures exist?

a. What steps should researchers take to protect data subjects or against societal-level harms prior to the dissemination of research outputs (publications, presentation slides, data visualization, datasets, AI/ML models, etc.)?

They should ensure that there is no way that individuals would be identified if others attempt to identify data subjects associated with particular insights or data. This should

be a consideration in reviewing and publishing processes as well. This includes carefully considering how to share the results with the communities that they took the data from. For instance, if a research team posted their results on a forum to get feedback from the community, that forum would be readily identifiable as the data source and thus easier to de-identify participants.

b. Under what circumstances is it appropriate for an online service provider or data intermediary to have access to or review third-party research papers before they are submitted for publication? Are there circumstances where pre-publication review is inappropriate?

It would be appropriate if there are concerns about the risks or harms that the work could expose their users/clients to.

It would be inappropriate if it limits freedom of academic expression.

c. Reproducibility can help promote trust in research. What factors do/should researchers consider when deciding when/how to delete, store, share, or archive pervasive data?

Reproducibility and the correlation of trust are norms in some disciplines, but not others. For example, in ethnographic methodology, reproducibility is not a benchmark of trust, but instead, reflexivity on the research process, transparency about access, and interpretation are useful contributions to ensure rigorous work (Pool, 2017). Adopting such norms could protect people and their data, and ensure researchers truly consider their need for using such data. Transparency over the researcher's personal storage of pervasive data with the potential for audit is also a possible consideration.

11. What existing ethical frameworks, such as those from professional organizations or government agencies, should be considered when drafting national-level ethical guidelines for research with pervasive data?

A number of ethical frameworks, including the U.S. Federal Data Ethics Framework (Federal Data Strategy Team), American Statistical Association (ASA) Ethical Guidelines for Statistical Practice (Committee on Professional Ethics of the American Statistical Association, 2022), Society for Research in Child Development (SRCD) Ethical Principles and Standards for Developmental Scientists (Society for Research in Child Development, 2021), and the Association of Social Anthropologists (ASA UK) of the UK Ethical Guidelines for Good Research

Practice (Association of Social Anthropologists of the UK) provide informed guidelines for general practices with research data that could be applied to drafting national-level ethical guidelines for research with pervasive data.

a. To what extent do existing frameworks apply to the collection and use of pervasive data?

All of the existing frameworks listed above do not explicitly mention pervasive data. However, given the similarities between pervasive data and other forms such as interview, survey, and measurement data, these frameworks are still highly relevant. For example, the Association of Social Anthropologists of the UK outlines measures such as withholding data from vulnerable or marginalized populations from being published, and nearly all of the aforementioned frameworks discuss best practices for informed consent, data security, and avoiding unnecessary risks to subjects.

b. What modifications of existing frameworks might be necessary to ensure that those frameworks are applicable to the needs of research with pervasive data?

To ensure existing ethical frameworks are applicable to research with pervasive data, modifications should address the unique challenges posed by its scale, granularity, and ubiquity. Frameworks like the U.S. Federal Data Ethics Framework and ASA guidelines should emphasize contextual integrity, dynamic consent, and bias detection while promoting interdisciplinary accountability and algorithmic transparency. SRCD principles must address children's data in pervasive digital contexts, ensuring developmental appropriateness, re-consent mechanisms, and protections for marginalized groups.

Given the rising interest in applying machine learning techniques to analyzing pervasive data, existing frameworks must also account for such nuances. These include the potential for amplified biases due to unrepresentative or contextually inappropriate datasets, privacy risks from de-anonymization, and the ethical challenges of using data collected without explicit consent. Additionally, the cultural, social, and geographic contexts of the data must be respected to avoid misinterpretation or harm. Finally, transparency in model development, explainability of results, and alignment with ethical principles are critical to ensure responsible and equitable use of pervasive data.

13. What structured processes (questionnaires, rubrics, assessment frameworks) could be used to determine which techniques should be used to mitigate risks to data subjects and society in research that relies on pervasive data?

In addition to frameworks from professional organizations or government agencies, researchers have developed frameworks and perspectives for guiding processes that mitigate risks to data subjects, such as: Citron's intimate privacy lens (Citron, 2022) and Nissenbaum's contextual integrity framework (Nissenbaum, 2004).

14. How should ethical guidelines take into account future technological advances around research with pervasive data?

In order to take future technological advances into account, the guidelines should do the following:

- Be reviewed every 6 -12 months (could consider establishing a working group that is responsible for this process) because as technology evolves, so should these guidelines.
- Rely on the principle of precedence - learn from past decisions to inform future ones (which may be different technology but similar ethical considerations).
- Require researchers to certify a "do no harm" statement that acknowledges that the capabilities of technology will evolve, but they should never use the data in a way that could negatively impact or be used against the person who supplied it (even if they consented for it to be used for future research needs). It could include a section on what a researcher can and cannot do with pervasive data, and shift the responsibility for unforeseen negative consequences onto the researcher rather than the data subject.
- Account for the perspectives and expectations of data subjects, with attention to the unique contexts and identities implicated in data collection, analysis, etc. (e.g. if the degree of perceived sensitivity of data type A is high for group B, then perhaps said data should not be collected from group B to begin with).
- Update the understanding of what situations require what kind of privacy protection, accounting for what data types are sensitive to data subjects (e.g. data about affect/emotions that is increasingly relevant in pervasive data collection/inferences/use, and are considered to be "sensitive" by data subjects (Andalibi & Buss, 2020).
- Consider how federal, state, and international data privacy regulations align and conflict and communicate these limitations to researchers so they can consider how to apply them to their respective projects.

Additionally, researchers should seek out a new way for people to consent in a way that reacts to how their data might be used in the future (e.g., people contributing DNA/genetic data to

determine their ethnicity didn't know how technology would evolve and then learn about siblings they did not know existed). An example of work that is being done in the field to consider future use of data and what that might mean for consent norms is "Guidance for ensuring fair and ethical broad consent for future use. A scoping review protocol" (Maxwell et al., 2021). "Informed consent" may no longer be possible in many contexts, so exploring ways to enable people to have control over their pervasive data and privacy is of utmost importance.

References

Andalibi, N., & Buss, J. (2020, April). The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks. *CHI 2020*. https://par.nsf.gov/servlets/purl/10437787

Association of Social Anthropologists of the UK. (n.d.). *Ethical Guidelines for Good Research Practice*. ASA Ethics. https://www.theasa.org/ethics/guidelines.shtml

Citron, D. K. (2022). *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age*. WW Norton.

Committee on Professional Ethics of the American Statistical Association. (2022, February 1). *Ethical Guidelines for Statistical Practice*. American Statistical Association. https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice

Davies, H. (2022). Reshaping the review of consent so we might improve participant choice. *Research Ethics*, *18*(1), 3-12. https://journals.sagepub.com/doi/10.1177/17470161211043703

Federal Data Strategy Team. (n.d.). *Federal Data Strategy: Data Ethics Framework*. Resources.data.gov. https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf

Markham, A. (2012, January 10). Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society*, *15*(3), 334–353. https://doi.org/10.1080/1369118X.2011.641993

Maxwell, L., Gilyan, R., Chavan, S. A., Merson, L., Saxena, A., & Terry, R. (2021, February 11). Guidance for ensuring fair and ethical broad consent for future use. A scoping review protocol. *F1000Research*, *10*, 102. https://doi.org/10.12688/f1000research.51312.1

McDonald, A. M., & Cranor, L. F. (n.d.). The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society 2008 Privacy Year in Review issue*. https://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf

National Academy of Medicine. (2020). Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust. *The National Academies Press*. https://doi.org/10.17226/27110

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*(1). https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/

Pool, R. (2017, July 28). The verification of ethnographic data. *Ethnography*, *18*(3), 281-286. Sage Journals. https://doi.org/10.1177/1466138117723936

Razi, A., Alsoubai, A., Kim, S., Naher, N., Ali, S., Stringhini, G., De Choudhury, M., & Wisniewski, P. J. (2022, April 28). Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts)*, *39*, 1-9. Association for Computing Machinery Digital Library. https://doi.org/10.1145/3491101.3503569

Society for Research in Child Development. (2021, March 28). *Ethical Principles and Standards for Developmental Scientists*. Society for Research in Child Development. https://www.srcd.org/about-us/ethical-principles-and-standards-developmental-scientists

Wei, M., Consolvo, S., Kelley, P. G., Kohno, T., Roesner, F., & Thomas, K. (2023, April 19). “There’s so much responsibility on users right now:” Expert Advice for Staying Safer From Hate and Harassment. *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, *190*, 1-17. Association for Computing Machinery Digital Library. https://doi.org/10.1145/3544548.3581229

Wisniewski, P., Siek, K., Butler, K., Allen, G., Shi, W., & Parashar, M. (2025, January 15). *Prioritizing Computing Research to Empower and Protect Vulnerable Populations*. Computing Research Association. https://cra.org/wp-content/uploads/2025/01/2024-2025-CRA-Quad-Paper_-Prioritizing-Computing-Research-to-Empower-and-Protect-Vulnerable-Populations.pdf