CENTER FOR DEMOCRACY & TECHNOLOGY

*January 15, 2025*

National Telecommunications and Information Administration
Department of Commerce
1401 Constitution Avenue, NW, Room 4725
Washington, DC 20230

Re: Docket No. 241204-0309, Ethical Guidelines for Research Using Pervasive Data

The Center for Democracy & Technology (CDT) respectfully submits these comments in response to the National Telecommunications and Information Administration's (NTIA) request for comments (RFC) regarding ethical guidelines for research using pervasive data. CDT is a nonprofit 501(c)(3) organization fighting to advance civil rights and civil liberties in the digital age for all users. Among our priorities, CDT advocates for greater privacy protections, protecting users' rights to access information freely, and ensuring online services enable individuals to exercise choice and control over their online experience. CDT also conducts and supports original, objective research to inform current policy regarding digital civil rights and liberties and advocates for ethical and privacy-protective researcher access to data to ensure technology governance is informed by real world data.

For purposes of responding to the RFC, these comments adopt the RFC's use of the terms "pervasive data"[1] and "online services"[2] and uses "research" to mean the systematic analysis of data to build and disseminate knowledge by independent researchers who operate outside of an online service provider and are often affiliated with an academic or non-profit institution, consistent with the RFC's focus on such data flows.[3]

[1] National Telecommunications and Information Administration (NTIA), Department of Commerce, Notice, Request for Public Comments, Ethical Guidelines for Research Using Pervasive Data, at: https://www.federalregister.gov/documents/2024/12/11/2024-29064/ethical-guidelines-for-research-using-pervasive-data at n.1. ("The term pervasive data is intended to mean data about people—user-contributed, observed, derived, or inferred—collected through online services regardless of the extent to which the data is publicly available, is aggregated, or could lead to the identification of an individual. Pervasive data may include text, images, videos, biometric information, information about a data subject's behavior (purchases, financial standing, media consumption, search history, medical conditions, location, etc.), and other information that makes up a person's digital footprint.").
[2] *Id.* at n.1. ("Online services may include a wide range of information technologies throughout the technology stack/technical infrastructure, including but not limited to web-based monitoring tools, content delivery networks, blockchain technology, digital labor platforms, education technology, Internet of Things devices, connected cars, wearable devices, mobile sensors, data brokers, streaming services, search engines, online marketplaces, social media platforms, and AI systems.").
[3] *Id.* at n.3.

CENTER FOR DEMOCRACY & TECHNOLOGY

## I. Research with Pervasive Data is Critical to User-Centered Online Service Governance

Researcher access to pervasive data on online services allows researchers to examine critical questions of public concern regarding the governance of online services, including the efficacy of content moderation, the spread of mis- and disinformation, the effects of online advertising, and AI use cases.[4] While the translation of research insights into actionable policy is neither simple nor straightforward,[5] providing researchers access to pervasive data about online services can lead to key insights to support evidence-based policymaking and foster online service accountability.

At the same time, researcher access to pervasive data on online services can pose privacy and legal risks to users, and research that fails to meet appropriate ethical standards risks undermining public trust in important technology research.[6] Moreover, while some research may constitute human subjects research and be subject to certain U.S. regulations ensuring compliance with certain privacy and ethical standards,[7] much of it is not - suggesting the need for ethical guidance detailing how research with pervasive data from online platforms can proceed to both inform online service governance and protect users' privacy, legal, and other rights.

Research with pervasive data is currently informed by ethical guidance from multiple sources, including the Menlo Report (supplementing the Belmont Principles and applying them to network and security research)[8] and guidance from the Association of Internet Researchers,[9] American Statistical Association, [10] and others, which provide helpful guidance regarding ethical considerations in research with pervasive data from online services. Uniform and generally applicable guidance for this kind of research

---

[4] *See* C. Vogus, Independent Researcher Access to Social Media Data: Comparing Legislative Proposals (2022) https://cdt.org/insights/independent-researcher-access-to-social-media-data-comparing-legislative-proposals/; G. Nicholas, Grounding AI Policy: Towards Researcher Access to AI Use Data (2023) https://cdt.org/wp-content/uploads/2024/08/2024-08-12-CDT-Research-Grounding-AI-Policy-report-final.pdf.

[5] M. Luria & A. Bhatia, Insights from a Child Safety Online Symposium: Bridging Research and Policy (2023) https://cdt.org/insights/insights-from-a-child-safety-online-symposium-bridging-research-and-policy/.

[6] C. Fiesler, et al. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research. Proceedings of the ACM on Human-Computer Interaction, 8(GROUP), 1-33 (2024) https://dl.acm.org/doi/pdf/10.1145/3633070, *citing* B. Hallinan, et al. Unexpected Expectations: Public Reaction to the Facebook Emotional Contagion Study. New Media & Society 22, 6 (2020) https://journals.sagepub.com/doi/10.1177/1461444819876944; J. Metcalf & K. Crawford. 'Where are human subjects in big data research? The emerging ethics divide.' Big Data & Society, 3(1) (2016) https://doi.org/10.1177/2053951716650211; and Michael Zimmer. 2016. OkCupid Study Reveals the Perils of Big-Data Science. WIRED (2016) https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/.

[7] 45 C.F.R. § 46.

[8] The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research (2012) https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf (hereinafter Menlo Report).

[9] Association of Internet Researchers, Internet Research: Ethical Guidelines 3.0 (2020) https://aoir.org/reports/ethics3.pdf (hereinafter AOIR Guidelines).

[10] Ethical Guidelines for Statistical Practice, American Statistical Association (2022) https://www.amstat.org/docs/default-source/amstat-documents/ethicalguidelines.pdf.

akin to the Belmont Principles, however, has yet to emerge. In response to the ongoing need for access to pervasive data on social media, the emerging need for access to pervasive data from AI systems and other online services, and the overarching need for researchers to have access to pervasive data in ways that are ethical and privacy-protective, CDT provides the following suggestions when considering uniform principles to guide such research.

## II. Application & Limitations of Existing Human Subjects Research Ethical Standards to Research with Pervasive Data

As noted by the RFC, ethical standards for research involving human subjects in the United States are rooted in the 1979 Belmont Report. While developed primarily with biomedical research and prominent biomedical ethics lapses in mind,[11] its principles provide a basic framework for understanding ethical obligations owed to subjects in research involving human subjects, but have important limitations that should be considered when applied to research with pervasive data.[12]

The Belmont Report identifies three basic principles ("Belmont Principles") for application in research: respect for persons, beneficence, and justice.[13] Respect for persons requires that researchers acknowledge autonomy and protect those with diminished autonomy in human subjects research. Beneficence requires that human subjects are treated in an ethical manner by respecting their decisions, protecting them from harm, and by making efforts to secure their well-being.[14] The principle of justice asks researchers to consider how the benefits and burdens of research are distributed.[15] Most commonly, these principles are demonstrated in research with pervasive data through harm minimization, obtaining informed consent, and protecting subject privacy and confidentiality.[16] Considerations of justice may also be demonstrated in the selection of research participants or in the distribution of research findings to both participants and the public.

Taken together, the Belmont Principles provide a rights-driven and user-focused basis upon which to build consensus ethical guidelines for research with pervasive data. At the same time, there are limitations to the wholesale and exclusive application of the Belmont Principles to research with pervasive data. The first limitation is both legal and practical. In the United States, the Belmont Principles are operationalized through the Common Rule, which applies to research conducted or funded by the federal government and involving human subjects.[17] Under the Common Rule, Institutional Review Boards (IRBs) are charged with evaluating prospective research projects to ensure

[11] Office of the Secretary, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1978) https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf (hereinafter "Belmont Report").
[12] AOIR Guidelines, *supra* note 9, at 4.
[13] Belmont Report, supra note 11.
[14] *Id.*
[15] *Id.*
[16] Fiesler, et al., *supra* note 6.
[17] 45 C.F.R. § 46.

compliance with human subjects research regulations, consistency with ethical standards and institutional policies, and adequate protection of human subjects.

Much research with pervasive data, however, does not constitute research with human subjects or is not subject to the Common Rule.[18] In most relevant part, the Common Rule defines a "human subject" as "a living individual about whom an investigator conducting research . . . obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens."[19] Information is considered "identifiable" if "the identity of the subject is or may readily be ascertained by the investigator or associated with the information."[20] To the extent that data is deidentified or available publicly, then, many observational studies utilizing pervasive data may not involve "human subjects," meaning that they are not subject to the Common Rule or IRB review. Moreover, independent researchers - particularly those at nonprofits or not affiliated with a university that receives federal funding - may not be subject to research ethics oversight, as the Common Rule only applies to federally funded research. A large proportion of research with pervasive data, therefore, does not qualify for IRB review, and many research projects with pervasive data may never undergo any form of ethical review.[21]

A second limitation is the incompatibility of users' relationship with controllers of pervasive data with the appropriately high informed consent requirements that flow from the principle of respect for persons. Most users, when interacting with an online service, agree to or are subject to terms of service and privacy policies implemented by the online service detailing the rules for use of the service and how the company will handle data collected by the platform. Some online services expressly contemplate researcher access to data in their terms of service and privacy policies. For example, Meta's Terms of Service reference its Privacy Policy, which notes that Meta shares user data with independent researchers "to conduct research that advances scholarship and innovation, and to promote safety, security and integrity."[22] Users, therefore, may theoretically know that, by continuing to use an online service, they may be subject to research on their behavior. In reality, however, most users do not fully read or comprehend online services' terms of service or privacy policies, or may not fully understand the implications of the policies to which they have agreed or are subject to.[23] This problem is exacerbated when data is sourced from data brokers, with whom the user has no contractual relationship and may have no knowledge that the broker holds or has purchased their data.

---

[18] Metcalf & Crawford, *supra* note 6.
[19] 45 C.F.R. § 46.
[20] 45 C.F.R. § 46.102(f).
[21] Q. Waeiss, Understanding AI Research Ethics as a Collective Problem (2023) https://www.adalovelaceinstitute.org/blog/ai-research-ethics-collective-problem/.
[22] Meta, Terms of Service, at: https://www.facebook.com/terms.php/; Meta, Privacy Policy, at: https://www.facebook.com/privacy/policy.
[23] A. McDonald & L. Cranor, The Cost of Reading Privacy Policies." ISJLP 4 (2008) https://www.technologylawdispatch.com/wp-content/uploads/sites/54/2013/02/Cranor_Formatted_Final1.pdf; Pew Research Center, Americans' Attitudes and Experiences with Privacy Policies and Laws (2019) https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/ (noting that 9% of adults report fully reading a company's privacy policy before agreeing to its terms and conditions).

Even if users are aware that their data may be used for research by an online service, data broker, or independent researchers, this knowledge does not constitute informed consent to research. The Belmont Principles state clearly that "[r]espect for persons requires that subjects, to the degree that they are capable, be given the opportunity to choose what shall or shall not happen to them," and that for consent to be informed, it must be supported by adequate and specific information, presented in a way that allows subjects to comprehend what they are agreeing to, and be done voluntarily.[24] In reality, then, a significant portion of research with pervasive data – collected from users who may never read a platform's terms of service or privacy policy and who, in any event, are not told about the specific research projects in which their data will be used – cannot be understood to be conducted with the informed consent of users. To require informed consent for each research project would severely limit researcher access to data and upend valuable research programs that advance knowledge and improve user safety. Simply because informed consent may not always be practical, however, does not minimize researchers' obligations of respect for persons, beneficence, or justice. Rather, it emphasizes the need for researchers using pervasive data to take particular care to protect the privacy and dignity of subjects whose data is the subject of their inquiry.

Such care is particularly warranted for researchers contemplating the use of pervasive data held by data brokers. The data broker industry collects and sells vast quantities of user data, with significant implications for users' civil rights and liberties, and with insufficient oversight and accountability.[25] Poor security practices have resulted in the exposure of data from tens of millions to over a billion people's data, leading to financial, reputational, and other privacy harms.[26] Ethical guidelines for research using pervasive data should account for risks to subjects stemming from the use of data, like data held by data brokers, collected in ways that are inconsistent with respect for the rights and privacy interests of individuals.

A third limitation is inherent to the Belmont Principles themselves, which focus primarily on the ethical relationship between researcher and subject, rather than the research's broader societal impact. To the extent that research with pervasive data receives ethical review, it is most often done through an IRB. IRBs, however, are expressly prohibited from considering broader societal or political impacts of research, instead focusing primarily on the interests of subjects.[27] For example, a research project attempting to measure the effect of artificial intelligence on engagement in political discourse by posting AI-generated images to a forum and measuring user responses may not involve human subjects

[24] Belmont Report, *supra* note 11.
[25] R. Shetty, CDT Comments to CFPB Lay Out Data Broker Harms That Should Be Held Accountable (2023) https://cdt.org/insights/cdt-comments-to-cfpb-lay-out-data-broker-harms-that-should-be-held-accountable/; C. Shenkman, et al., Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers (2021) https://cdt.org/wp-content/uploads/2021/12/2021-12-08-Legal-Loopholes-and-Data-for-Dollars-Report-final.pdf.
[26] Shetty, *supra* note 25.
[27] 45 C.F.R. § 46.111(a)(2) ("The IRB should not consider possible long-range effects of applying knowledge gained in the research (*e.g.,* the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility.").

but could have effects on election outcomes, raising significant ethical considerations that may not be addressed by an IRB. U.S. regulations appropriately prioritize subject protection and research operations in examining research ethics under the Common Rule. At the same time, doing so excludes from standardized ethical review the broader societal and ethical considerations that flow from research with pervasive data.[28] Given the inherent limitations of IRB review, institutions and others have begun processes and created guidance for considering the broader implications of research with pervasive data.

Some institutions, such as Stanford University through its Ethics and Society Review, have begun offering or requiring broader ethical review to assess how research may pose risks to those beyond human subjects.[29] The 2012 Menlo Report, Ethical Principles Guiding Information and Communication Technology Research, proposes a fourth principle to supplement the Belmont Principles in recognition of the Belmont Principles' limited scope - namely respect for law and public interest. While arguably implicit in the Belmont Principles, the principle of respect for law and public interest requires researchers to engage in legal due diligence, be transparent in methods and results, and be accountable for actions.[30] Legal and public interest considerations are of unique importance to research involving pervasive data in the United States, given the absence of comprehensive legal privacy protections and the potential for research with pervasive data to have consequences well beyond the boundaries of its own research.[31] Respect for law and public interest requires researchers to ensure compliance with applicable laws, such as the Computer Fraud and Abuse Act and copyright, and to be transparent in how and why data is to be collected, utilized, and how research results will be utilized to avoid undermining trust.[32]

## III. Points to Consider in Applying Ethical Guidance for Research Using Pervasive Data

Given the limitations of the Belmont Principles and the need to advance ethical research with pervasive data to develop knowledge and improve users' experiences, government and civil society should collaborate to explore the establishment of ethical norms that allow for rigorous research with pervasive data that serves the interests of users and society. Four points to consider in doing so follow.

### A. Ethical Guidelines Should be Applied Based on the Context of the Pervasive Data

The definition of pervasive data on online services covers an extraordinary amount of data with differing characteristics. The ethics of research with public-facing social media posts, for example, differs significantly from data as private or sensitive as geolocation data collected by connected cars or chat logs between a user and their AI intimate companion. To the extent that a uniform set of ethical principles for research with data this diverse is possible, at a minimum, any such principles should make

[28] Waeiss, *supra* note 21.
[29] Ethics & Society Review, Stanford University, https://casbs.stanford.edu/ethics-society-review-stanford-university.
[30] Menlo Report, *supra* note 8.
[31] *Id.*
[32] *Id.* at 16.

clear that their application will differ by data type, the unique aspects of each type of data, its provenance, and the associated risks to users subject to the research.

Consider one type of data within the definition of "pervasive data" - metadata about the use of social media to publicly post information about abortion and travel to Oregon by Idaho- and Oregon-based users. Research to study the effect of Idaho's Child Custody Protection Act, which prohibits the transportation of minors across state lines to obtain an abortion without parental consent,[33] would pose numerous challenging ethical questions. Ethical principles of respect for persons, beneficence, and justice suggest that in many cases users should be able to decline consent to participate in such research, potentially including circumstances in which the data has been deidentified given the risk of re-identification and for such data to be targeted for collection by Idaho law enforcement. While the same basic ethical principles that apply to this kind of research may also apply to research regarding, for example, consumer habits relating to the use of smart vacuums, the consequences of unauthorized disclosure or other mishandling of the former as compared to the latter suggest that the unique aspects of each kind of pervasive data should be considered in ethical analysis, and specific guidelines for uniquely sensitive subsets of data may be appropriate.

B. Differences in Power Between Services, Researchers, and Users

As appropriately noted by the RFC, power differences exist along at least three axes – between researchers and data subjects, online service providers and data subjects, and service providers and researchers – that create unique risks and potential for harm for users and researchers alike. Online services hold significant power to limit or prevent research by either denying access to data or retaliating against researchers. Providers may also target and undermine research by alleging terms of service violations for research that is unflattering to the service provider.[34] As such, independent researchers are frequently at legal or reputational risk for conducting their research, requiring both defensive and proactive litigation to pursue their aims.[35]

Online services also hold significant power over users and their data - subjecting them to terms of service and privacy policies that may expose their data to research in ways they never intended. For many types of data, users are dependent on online services to protect their privacy and intervene if their data is misused within the research context. When platforms are resistant to external research and accountability, then, user privacy interests can be co-opted by platforms to block appropriate access to pervasive data to perform research. Users are also subject to the power of researchers themselves who have the power to use pervasive data from online services to advance their research

[33] Idaho Code § 18-623.
[34] Brief of Amici Curiae Coalition for Independent Technology Research and Integrity Institute in X Corp. v. Center for Countering Digital Hate (2024) https://independenttechresearch.org/wp-content/uploads/2024/09/2024.09.27-Dkt.-43.1-CITR-and-Integrity-Institute-Amicus-Brief.pdf.
[35] *See, e.g.*, Knight First Amendment Institute at Columbia University, Zuckerman v. Meta Platforms (2024) https://knightcolumbia.org/cases/zuckerman-v-meta-platforms-inc.

objectives in ways that, most often, users never consent to, particularly in the case of data brokers, with whom a user has no contractual relationship. Providers' power over both users and researchers also limits how researchers can fulfill their ethical obligations to subjects. For example, since researchers may access data about users from a platform rather than directly from the users themselves, researchers are often not in a position to obtain informed consent to research and instead may be dependent on assurances made by the provider and their terms of service to understand what research may be done with the data.

Traditional research ethics acknowledge the imbalance of power between subject and researcher by prioritizing respect for persons, beneficence, and justice to both allow for the development and dissemination of knowledge and respect human rights. While the ethics of research with pervasive data should not be ignorant to the power differences between researchers and platforms, those differences do not limit or modify researchers' ethical obligations to subjects.

### C. Legal Risks and Rights of Users from Researcher Access to Pervasive Data

While access to pervasive data is important to foster transparency into the operation of online services through research, pervasive data is also a rich source of information for law enforcement agencies.[36] Geolocation data, employment data, reproductive and gender health related data, and other pervasive data across online services can reveal detailed information that law enforcement may use to surveil or otherwise invade the privacy of users, even leading to prosecution. The ethics of research with pervasive data should consider users' legal rights and risks when research with pervasive data could result in demands for access from law enforcement.

The Fourth Amendment protects people in the United States against unreasonable searches and seizures, requiring law enforcement to obtain a warrant in most cases. Warrants are based on "probable cause" and require court approval, while subpoenas can be issued without court approval and without probable cause, only needing to be relevant to an investigation. Whether the government needs a warrant to compel access to pervasive data shared with researchers depends on whether the courts deem accessing that data to constitute a "search." Courts apply a two-part test to determine if a search occurs: first, whether the individual has a subjective expectation of privacy in the data, and second, whether that expectation is one that society recognizes as reasonable. Publicly available data, such as public social media posts, generally does not meet this test, meaning law enforcement can access this data without a warrant. Recent Supreme Court decisions like *United States v. Jones* and *Carpenter v. United States*, however, have raised questions about whether pervasive collection of public data, such as long-term monitoring of social media posts, might violate a user's reasonable expectation

[36] C. Vogus, Defending Data: Privacy Protection, Independent Researchers, and Access to Social Media Data in the US and EU (2023) https://cdt.org/wp-content/uploads/2023/01/2023-01-23-CDT-Defending-Data-Independent-Researcher-Access-to-Data-report-final.pdf.

of privacy, particularly when sensitive information, like location data, is collected and can be the basis for inferences about private activities.[37]

Fourth Amendment protections are generally stronger for private information, but are complicated by the third-party doctrine. Under the third-party doctrine, "a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties."[38] As such, when an individual voluntarily shares information with a third party (like a social media platform), they may forfeit their reasonable expectation of privacy in that data, allowing the government to access it without a warrant. Courts are divided on how this doctrine applies to modern, sensitive data shared with online services, and, therefore, to independent researchers who access that data.[39] Some rulings suggest that users might retain an expectation of privacy in some forms of sensitive data even when shared with third parties, while others suggest that users may lose privacy protections when data is shared with researchers, especially if they are aware of the data sharing.[40]

The Stored Communications Act (SCA), part of the Electronic Communications Privacy Act of 1986, regulates government access to stored electronic communications. The SCA generally applies to providers of "electronic communications services" and "remote computing services," such as social media and other online platforms. It requires law enforcement to meet specific standards—ranging from subpoenas to warrants—depending on the type of information requested. The SCA, however, does not apply to researchers who receive pervasive data.[41] If social media companies share pervasive data with researchers, and law enforcement seeks that data from the researchers rather than the platform, the SCA's protections may no longer apply, potentially making it easier for law enforcement to compel access to the data using subpoenas instead of warrants. Moreover, the voluntary disclosure rules of the SCA[42] may also be inapplicable to researchers, enabling them to voluntarily share this information with law enforcement even though it could not be volunteered by the platform from which it was obtained by the researcher.

Ethical standards for research with pervasive data should take into account the risk that researcher access to data may make it easier for law enforcement to access the same information. To minimize the risk of legal harm to users, ethical guidelines should emphasize good data practices (such as retention limits on pervasive data), prohibitions or restrictions on voluntary disclosures to the government, and access practices that maintain as many protections for users as possible under the SCA and the Fourth Amendment.

### D. Considerations for Research on Pervasive Data Pertaining to Children

[37] *Id.*
[38] Smith v. Maryland, 442 U.S. 735, 743-44 (1979).
[39] Vogus, *supra* note 36.
[40] For a full discussion of the application of the third-party doctrine to data transferred for independent research and recent case law developments, please see Vogus, *supra* note 36.
[41] *Id.*
[42] 18 U.S.C. § 1702.

Children are at heightened risk from research on pervasive data due to their vulnerability and potentially more limited understanding of research, data, and data-related implications. As with research on pervasive data pertaining to adults, data that is collected and used for research may result in disclosure of children's personal data if not used responsibly.[43] Research has also shown that youth use online services and platforms as a means to test and consider new identities.[44] For some, such as those exploring aspects of their identity like their gender and sexuality, such exposure could risk revealing their activities and interactions still in their exploration stage, and causing tangible harm to their wellbeing. Pervasive data about children collected and used today could also have unforeseen consequences later in children's lives, potentially affecting their career path, reputation, and possible opportunities. This is especially true given youth's potentially lower and fragmented awareness of data privacy,[45] and the objective difficulty to assess how data may be used in the future. Youth are also more likely to have their perspectives and identities change and evolve over time,[46] making collected data potentially unrepresentative of their future character and beliefs.

Protecting the rights of children and youth in research involving pervasive data builds on general research ethical principles and best practices, but requires additional scrutiny and potentially higher standards. Researchers must navigate challenges of collecting and using youth pervasive data, balancing the potential benefits of research with the imperative to uphold the privacy and data rights of young research subjects. In light of children's unique vulnerability, ethical guidelines for research involving pervasive data on children should emphasize rigorous data minimization, deidentification, and data destruction practices to protect privacy rights and reduce risks of misuse or unauthorized access. Ethical guidelines should also emphasize the importance of clear communication to both children and their guardians regarding how data will be collected, used, and shared in research. Researchers should also prioritize the input of children, youth, and caregivers to ensure the representation of their perspectives in designing research studies, providing them with agency and a voice.

***

CDT appreciates the opportunity to provide input on the importance and ethics of independent research with pervasive data. For additional information, please contact Becca Branum at bbranum@cdt.org.

[43] Collection and use of personal data of known children, or on child-directed online services, would be subject to the Children's Online Privacy Protection Act, which would require parental consent for research uses of children's data. COPPA, however, covers only a portion of children's online data, and therefore the concerns around pervasive data collection of children's data still apply.

[44] D. Boyd, It's Complicated: The Social Lives of Networked Teens. Yale University Press, 2014.

[45] N. Santer, et al., Early Adolescents' Perspectives on Digital Privacy (2021) https://wip.mitpress.mit.edu/pub/early-adolescents-perspectives-on-digital-privacy/release/1.

[46] M. Luria & N. Foulds, Hashtag-Forget: Using Social Media Ephemerality to Support Evolving Identities (2021) https://dl.acm.org/doi/fullHtml/10.1145/3411763.3451734.