April 1, 2024

To: Director Shalanda D. Young
The Office of Management and Budget
725 17th St. NW
Washington, DC 20503

Submitted via regulations.gov

The Center for Democracy & Technology (CDT) is pleased to submit this comment on the value of and potential improvements to Privacy Impact Assessments (PIAs) from federal government agencies. CDT is a nonprofit 501(c)(3) organization that works to advance civil rights and civil liberties in the digital age, including privacy. CDT is supportive of PIAs, both as an analytical tool for risk management and mitigation and as a transparency mechanism for agencies' use of technologies.

However, PIAs should be improved and expanded to meet the needs of a more expansive technology ecosystem. For instance, taking a data minimization approach to collection and storage where possible can help to mitigate privacy threats, and OMB guidance should include analysis of an agency's data minimization procedures as part of a PIA. In addition, the categories of data currently required to be reviewed as part of a PIA are not sufficiently robust given the inferential and reidentification capabilities of technology. Seemingly non-identifying data points may be combined to identify individuals in some cases, and PIAs should account for this possibility. Additionally, the utility of a PIA as a transparency tool outlasts the lifespan of the system itself, but many agencies do not maintain historical PIAs for public access. Doing so would strengthen the transparency function of PIAs (and may also provide a library of risks and mitigations that would be useful to other agencies and system deployers).

Beyond improvements to PIAs themselves, OMB could make improvements to the broader transparency and risk mitigation frameworks that PIAs are part of. Given their longevity in supporting transparency around federal data and technology efforts, PIAs offer a relatively consistent and well-used framework by which agencies assess and publicize their technology use cases. However, there is no centralized repository or standard format for PIAs, which makes it difficult to compare PIAs across agencies or do a robust search for technologies (e.g. cybersecurity, AI, digital identity management) used by the federal government writ large.

This ecosystem will be further expanded and complicated by the addition of the federal inventory of AI uses[1] (which are even less consistent and thorough than PIAs[2]) and related analyses like Algorithmic Impact Assessments (AIAs). Therefore, OMB should focus on building

[1] AI.gov, Federal AI Use Case Inventories (Accessed Apr. 1, 2024) https://ai.gov/ai-use-cases/.
[2] US Government Accountability Office, Artificial Intelligence: Agencies Have Begun Implementation but Need to Complete Key Requirements (Dec. 12, 2023) https://www.gao.gov/products/gao-24-105980.

a framework and providing guidance that creates a cohesive system of assessments and analyses of government systems, allowing for holistic analysis and easily accessible documentation of the government's use of data and technology.

1: A wide range of privacy risks are associated with the creation, collection, use, processing, storage, maintenance, dissemination, disclosure, and disposal of personally identifiable information (PII). What improvements to OMB guidance on PIAs as analytical tools and notices to the public would assist agencies in identifying, addressing, and mitigating these risks

As noted above, PIAs are a valuable transparency and analytical tool. However, OMB should provide additional guidance to maximize their utility:

- *Assessment of data minimization practices.* As noted in the request for comment, a wide variety of privacy harms and risks are associated with all parts of the data life cycle. Not only can particular uses (or misuses) of data cause a series of harms, such as reputational , psychological, physical, and economic harms,[3] but the likelihood that data will be misused or accessed without authorization makes mere collection of data quite harmful.[4] And even less private, or less "sensitive," data can cause harm because so often non-sensitive data can be used as a proxy for sensitive data.[5] Thus, collection of data on its own can cause a significantly increased risk of harm, and agencies should be vigilant in ensuring their systems embody the data minimization principle.[6]

  The government already has adopted a policy to minimize the PII it collects under OMB Circular A-130.[7] Agencies are also tasked with creating inventories of their information systems that collect and use PII.[8] In pursuit of these policies, OMB should update guidance document M-03-22 in the section under "Conducting a PIA" to include analysis of the agency's data minimization procedures and whether it can take further steps to minimize the collection, use, and retention of personal data.

- *Expand types of data subject to PIAs.* OMB should consider expanding the data and information covered by PIAs, particularly for AI-based systems. Given the ability of AI systems to glean information from datasets, they may end up discerning PII from a data set that does not facially contain such information (this is particularly true of data that

[3] Danielle Keats Citron & Daniel Solove, *Privacy Harms*, 102 B.U. L. Rev. 793 (2022), https://www.bu.edu/bulawreview/files/2022/04/CITRON-SOLOVE.pdf.

[4] Center for Democracy & Technology Comments on Commercial Surveillance ANPR (Nov. 21, 2022), https://cdt.org/wp-content/uploads/2022/11/CDT-Comments-to-FTC-on-ANPR-R111004.pdf; Justin Brookman & G.S. Hans, *Why Collection Matters: Surveillance as a De Facto Privacy Harm*, CDT (2013), https://cdt.org/wp-content/uploads/2018/08/September-2013-Brookman-Hans-Why-Collection-Matters.pdf

[5] Daniel Solove, *Data Is What Data Does*, 118 Northwestern U. L. Rev. 1081 (2024), https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2923&context=faculty_publications.

[6] *See* OMB Circular A-130, Appendix II, at II-3, https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/circulars/A130/a130revised.pdf.

[7] *Id*. at section 5(f)(1)(e) (Agencies shall, "[t]o the extent reasonably practicable, ensure that PII is accurate, relevant, timely, and complete, and reduce all PII to the minimum necessary for the proper performance of authorized agency functions").

[8] *Id*. at section 5(a)(ii).

once contained PII but has been deidentified).[9] OMB should provide guidance to agencies about the types of data that are most likely to raise re-identification concerns, such as ZIP codes, email addresses and screen names, and demographic information.[10] Agencies should assess these non-identifying but high-risk data points to determine whether any combination of those points could ultimately identify individuals with some reasonable degree of certainty. This analysis should be included in PIAs, along with any mitigations set in place to avoid that identification (such as eliminating one of the points, adding statistical noise to the data, or other obfuscation approaches).

- *Posting historical PIAs.* OMB should issue guidance to agencies that directs them to continue to publicly post historical PIAs, even after the systems they were developed for are no longer active. This historical information is important for transparency as it can help advocates and communities understand how agencies have historically used their data. Additionally, the potential impacts of a system do not vanish the moment a system is decommissioned. Whether that is data that was breached and held until a later date, or a decision made by a now-decommissioned system that continues to exist beyond the lifespan of the original system, the repercussions of a system can echo on. Consequently, historical PIAs should remain available for an extended period of time.

- *Standardized format.* While PIAs contain standardized information, their formats and structures vary across agencies. This can make it challenging for a user who is pulling PIAs from multiple agencies to assess and compare them.[11] Requiring a standardized format would increase the usability of PIAs for readers.

- *Plain-language summary.* Many PIAs are not well designed for a member of the general public. They contain significant technical information, often relying on domain-specific terminology. While this is reasonable for the analytical role of a PIA, it is a shortcoming for a public notice. PIAs could be made more accessible to a much wider range of users by including a plain-language summary of the system being assessed and the key risks and mitigations.[12]

3: What guidance should OMB consider providing to agencies to help reduce any duplication that may arise in preparing PIAs along with other assessments focused on managing risks (e.g., security authorization packages or the AI impact assessments proposed in OMB's Draft

[9] Boris Lubarsky, Re-Identification of "Anonymized" Data, Georgetown Law Technology Review (April 2017) https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/.

[10] Rajesh Parthasarathy, Reidentification Risk Of Masked Datasets: Part II, Forbes Technology Council (Jul. 27, 2020) https://www.forbes.com/sites/forbestechcouncil/2020/07/27/reidentification-risk-of-masked-datasets-part-ii/?sh=38841d5f5419.

[11] See, for example, PIAs from the Department of Veterans' Affairs (VA) and the Department of Health and Human Services (HHS), respectively: https://department.va.gov/privacy/wp-content/uploads/sites/5/2023/05/FY22StaffSergeantGordonFoxSuicidePreventionPIA.pdf; https://www.hhs.gov/sites/default/files/ahrq-medical-expenditure-panel-survey-enclave.pdf.

[12] The VA PIA above, for example, includes a summary of the system and its intended purpose, but that summary contains terminology that would not be accessible to many readers from the general public.

Memorandum on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence) and to support these assessments' different functions?

The principal area in which OMB should focus its coordination and burden reduction efforts is between PIAs and the new AIAs, described in the Memorandum from OMB. AIAs are analyses of AI-based systems designed to assess the potential harms of such a system as well as document its intended uses and expected benefits.[13] The potential risks stemming from algorithmic systems will certainly include privacy considerations, meaning that the risks raised in a PIA are likely to be relevant to any system that requires an AIA as well. Additionally, AIAs require agencies to consider the data involved in a system, examining its quality, its reliability, and its relevance for the given system. Because AI-based systems generally require a significant amount of data, most systems that require AIAs will require a PIA as well.

For instance, collection and use of personally identifiable information (PII) will engender privacy risks (e.g., risk of a data breach due to storage of that information and risk of a generative AI system disclosing PII as an output to unauthorized parties), and the majority of AI-driven systems that implicate rights and safety risks will include PII and/or information that is derived from PII.

OMB should provide guidance about how to assess and document these shared risks that span PIAs and AIAs. For instance, the guidance should require agencies that note in their PIA process that their system collects biometric information to consider the equity implications of that collection in an AIA for the same system. To reduce burden, agencies should only have to document this once and have the other assessment reference the source analysis. This limits the likelihood that such shared risks will be missed, and may make it easier to identify certain risks across assessments.

Additional risks that span privacy and rights- and safety-impacting AI include:

- Biometric data raises privacy risks as well as equity risks that should be addressed in an AIA.
- Demographic data raises equity considerations in that it may result in systems “learning” and regurgitating stereotypes, and also may raise reidentification risks for small minority populations.
- Proxy data such as ZIP code information raises privacy risks that are exacerbated by AI-based systems with more computing power that make reidentification more likely.

[13] Office of Management and Budget, Memorandum for the Heads of Executive Departments and Agencies on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (Mar. 28, 2024) https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

5: What improvements to PIAs would help you better understand agencies' assessment of privacy impacts and risk mitigation strategies? (a.) What improvement(s) would you recommend to make it easier to find and access agencies' PIAs?

OMB should build off the existing framework it has created in centralizing agencies' AI inventories, including the improvements reflected in comments on the Proposed Memo on Agency Use of AI as well as in the final version of OMB's AI memo.[14] This would mean providing a centralized repository for PIAs from multiple agencies (particularly for common technology), in addition to AIAs. OMB should create a mechanism for users to receive alerts about updates and modifications to PIAs in the consolidated inventory and make the inventory accessible through programmatic means, like an API. This would allow for far more robust analysis of PIAs and government systems in general by advocates and the general public.

Additionally, OMB should develop a standard for summarizing PIAs to make them more immediately accessible to a wide range of readers. While the content of PIAs is largely standardized, the summaries often omit information that is relevant for making a transparent system. These *summaries* should include:

- The data the system collects
- The agency's stated goal for using the system
- The intended lifespan of the system
- List of the parties who will have access to the system

OMB must ensure agency submissions are complete and easily understood by the broader public. To maximize efficiency for OMB and agencies alike, this should be achieved through clear guidance. OMB could draw on guidance from Inspectors General for managing spending records,[15] including:

- Provide example language and structured questions
- Ensure that entries (and modifications) are dated
- Consider minimum length requirements for fields that warrant longer explanations, such as the AI Use Case Summary (clarifying that a 1-2 sentence submission is insufficient)
- Instruct agencies to avoid jargon and abbreviations[16]

[14] Dan Bateyko, Hannah Quay-de la Vallee, Ridhi Shetty, and Alexandra Reeve Givens, CDT Comments on OMB Draft Guidance for Agency Use of AI, Center for Democracy & Technology (Dec. 5, 2023) https://cdt.org/wp-content/uploads/2023/12/CDT-Comments-on-OMB-2023-0020-Draft-Guidance-for-Agency-Use-of-AI.pdf.

[15] U.S. Government Accountability Office, DATA Act: OIGs Reported That Quality of Agency-Submitted Data Varied, and Most Recommended Improvements | U.S. GAO, (2020), https://www.gao.gov/products/gao-20-540.

[16] See, e.g. Office of Inspector General, Dept. of Homeland Security, DHS Has Made Progress in Meeting DATA Act Requirements, But Challenges Remain (2020), https://www.oig.dhs.gov/reports/2020/dhs-has-made-progress-meeting-data-act-requirements-challenges-remain /oig-20-62-aug20 ("DHS did not use plain English language when reporting award descriptions as required. Officials used shorthand descriptions, abbreviations, or terminology that could only be understood by officials at the Department or component that made the award.").

- Develop automated checks and controls in the submission system, for example via required fields.

These frameworks will help ensure PIAs (as well as other assessments) serve their function as a transparency tool. To further enhance this transparency framework, OMB should ensure that in their centralized repository of PIAs and AI systems, all PIAs and AIAs of a given system are easily accessible from the same place. For instance, a spreadsheet-style collection of systems should include PIAs and AIAs on a single row dedicated to a given system.

6: How can agencies increase awareness of PIAs among stakeholders?

As noted above, a centralized repository of PIAs would make them more accessible to users, particularly those who are interested in assessments from multiple agencies. Plain-language summaries would also increase the usability of PIAs by individuals as well as community and advocacy organizations. These usability improvements should come alongside efforts to increase awareness as part of a concerted effort to garner more engagement with PIAs.

One approach to increasing awareness of PIAs, as well as improving the assessments, is to include stakeholders, particularly impacted communities, as part of the PIA process and, if an AIA is conducted, that process as well. Community engagement can be a challenging process but, as noted by OIRA, it offers critical insights and information for both risk management and effective design and deployment of systems.[17] Impacted communities will often have a unique perspective on the potential harms that can stem from use of their data or data about them.[18] Consequently, incorporating them into the assessment process improves the likelihood that agencies will discover all relevant risks and also may provide insight into the most effective mitigations for those risks.

7: AI and AI-enabled systems used by agencies can rely on data that include PII, and agencies may develop those systems or procure them from the private sector.

a. What privacy risks specific to the training, evaluation, or use of AI and AI-enabled systems (e.g., related to AI system inputs and outputs, including inferences and assumptions; obtaining consent to use the data involved in these activities; or AI-facilitated reidentification) should agencies consider when conducting PIAs?

AI systems often significantly complicate dataflows, meaning that tracing how a specific piece of data moves through an agency's systems can be challenging, if not impossible. Consequently, PIAs for AI-based systems should include any outputs of the system that might pull from the

[17] Office of Information and Regulatory Affairs, Broadening Public Participation and Community Engagement in the Regulatory Process (Jul. 19, 2023) https://www.whitehouse.gov/wp-content/uploads/2023/07/Broadening-Public-Participation-and-Community-Engagement-in-the-Regulatory-Process.pdf.

[18] Elizabeth Laird and Hugh Grant-Chapman, Report – Sharing Student Data Across Public Sectors: Importance of Community Engagement to Support Responsible and Equitable Use, Center for Democracy & Technology (Dec 2. 2021) https://cdt.org/insights/report-sharing-student-data-across-public-sectors-importance-of-community-engagement-to-support-responsible-and-equitable-use/.

system's inputs in a way that could expose information. (For instance, a generative AI system could redisclose information it has learned from inputs). PIAs should account for these risks and explain any mitigations that will be put into place.

Additionally, as mentioned above, data minimization is a key privacy principle recognized by OMB that it should operationalize through privacy impact assessments. Minimization is particularly important in AI systems. The growth and success of AI systems has turbocharged data collection, directly counter to the data minimization principle. Often, AI systems are trained on unfathomably large datasets that include data that may or may not be useful to the system, and the datasets are generally not reviewed for privacy risks or violations. Thus, training AI systems creates more privacy risk as data is now collected and retained indefinitely for AI training and retraining purposes. Privacy impact assessments that cover AI systems should include, as mentioned above, analysis of data minimization efforts used in the AI system, including its training datasets and any alternative AI systems considered.

b. What guidance updates should OMB consider to improve how agencies address and mitigate the privacy risks that may be associated with their use of AI?

As mentioned above, training data should be treated with particular caution, even if it does not appear to contain PII. In particular, AI systems may be capable of reidentifying datasets that were deidentified. Therefore, OMB should update its guidance to analyze training data for high-risk proxy data types; those data types that, while not directly identifying, are at high risk for *becoming* identifying when combined with another data point. These high-risk data points would include categories like:

- Zip code,
- Email addresses and screen names, and
- Demographic data.

As noted above, because of the complexity of data flows, it may not always be possible to ensure that data that is input into AI systems will never appear in their output. Consequently, risk management will have to involve monitoring and assessments of systems throughout the course of their deployment. OMB should issue guidance on how to create a PIA that can analyze a system over time and capture evolving risks and mitigations. If the system is found to present privacy risks by redisclosing information, any updates to the system should be incorporated into the PIA. This means that PIAs will have to be living documents throughout the lifespan of the system. Agencies should maintain versions of the PIA to allow readers to see how the risk assessment and mitigations have evolved over time.

PIAs can be a valuable framework for both system and analysis, and we appreciate OMB's interest in ensuring that they are optimally useful, and remain effective even as the technology used by the government evolves over time.