RECOMMENDATION: Data Challenges and Privacy Protections for Safeguarding Civil Rights in Government

The National Artificial Intelligence Advisory Committee (NAIAC)

May 2024

NAIAC MEMBERS

Miriam Vogel, *Chair*
President and CEO of EqualAI

James Manyika, *Vice Chair*
Senior Vice President, Google, President for Research, Technology & Society

Amanda Ballantyne
Director of the AFL-CIO Technology Institute

Jack Clark
Co-founder of Anthropic

David Danks
Professor of Data Science and Philosophy at the University of California, San Diego

Victoria A. Espinel
President and CEO of BSA | The Software Alliance

Paula Goldman
Chief Ethical and Humane Use Officer at Salesforce

Susan Gonzales
Founder and CEO of AIandYou

Janet Haven
Executive Director of Data & Society

Daniel E. Ho
William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, and Professor of Computer Science (by courtesy), Senior Fellow, Stanford Institute for Human-Centered AI, Stanford University

Ayanna Howard
Dean of Engineering at The Ohio State University

Jon Kleinberg
Professor in the Departments of Computer Science and Information Science at Cornell University

Ramayya Krishnan
W. W. Cooper and Ruth F. Cooper Professor of Management Science and Information Systems, Carnegie Mellon University

Ashley Llorens
Vice President, Distinguished Scientist, and Managing Director at Microsoft Research

Haniyeh Mahmoudian
Global AI Ethicist at DataRobot, Inc.

Christina Montgomery
Chief Privacy & Trust Officer and Vice President at IBM

Liz O'Sullivan
CEO of Vera

Fred Oswald
Professor and Herbert S. Autrey Chair in Social Sciences, Rice University

Trooper Sanders
CEO of Benefits Data Trust

Navrina Singh
Founder and CEO of Credo AI

Swami Sivasubramanian
Vice President for Data and Machine Learning Services at Amazon Web Services

Keith Strier
Vice President for Worldwide AI Initiatives at NVIDIA

Reggie Townsend
Vice President of Data Ethics at SAS Institute

INTRODUCTION

An important concern with the rise of AI systems is the potential for exacerbating bias and algorithmic discrimination. Recent executive orders reflect the importance of ensuring the federal government's use of AI systems are consistent with broader policies to advance equity and protect against unlawful discrimination. For instance, Executive Order (EO) 13,985 on *Advancing Racial Equity and Support for Underserved Communities Through the Federal Government* explicitly requires federal government agencies to conduct assessments of the differential impact of federal policies and programs on demographic groups; EO 14,091 on *Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government* extends equity-related requirements for federal agencies to AI and automated systems;[1] and EO 14,110 on *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* reinforces the federal government's commitment to advancing racial equity through its AI policies and systems.[2] The importance of understanding the potential for bias is reflected as well in prior NAIAC recommendations[3] and much outside research.[4]

Central to mitigating algorithmic discrimination is understanding the impact of AI systems in government programs and services. EO 13,960 on *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government* highlights the importance of assessing the impact of AI systems and ensuring the use of AI systems is consistent with civil rights and existing anti-discrimination laws[5], with the Blueprint for an AI Bill of Rights, the NIST AI Risk Management Framework, and EO 14,091 reinforcing this idea and highlighting the importance of equity

---

[1] Sec. 4(b) of EO 14091 instructs "[w]hen designing, developing, acquiring, and using artificial intelligence and automated systems in the Federal Government, agencies shall do so, consistent with applicable law, in a manner that advances equity."

[2] Sec. 2(b) of 14110 instructs agencies to adhere to principles including that "Artificial Intelligence policies must be consistent with my Administration's dedication to advancing equity and civil rights. My Administration cannot — and will not — tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice....It is necessary to hold those developing and deploying AI accountable to standards that protect against unlawful discrimination and abuse, including in the justice system and the Federal Government. Only then can Americans trust AI to advance civil rights, civil liberties, equity, and justice for all."

[3] NAIAC Year 1 Report, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf.

[4] See Jennifer King, Daniel E. Ho, Arushi Gupta, Victor Wu, and Helen Webley-Brown. 2023. The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 492–505. https://doi.org/10.1145/3593013.3594015; de Souza Briggs, X. & McGahey, R., 2022. *Keeping promises while keeping score: Gauging the impacts of policy proposals on racial equity*, Brookings Institution. United States of America. https://www.brookings.edu/articles/keeping-score-measuring-the-impacts-of-policy-proposals-on-racial-equity/.

[5] Two of EO 13,960's "Principles for Use of AI in Government" touch upon this, including principle #1 ("Lawful and respectful of our Nation's values. Agencies shall design, develop, acquire, and use AI in a manner that exhibits due respect for our Nation's values and is consistent with the Constitution and all other applicable laws and policies, including those addressing privacy, civil rights, and civil liberties.") and principle #2 ("Purposeful and performance-driven. Agencies shall seek opportunities for designing, developing, acquiring, and using AI, where the benefits of doing so significantly outweigh the risks, and the risks can be assessed and managed.").

assessments to detect algorithmic bias.[6] The Office of Management and Budget's (OMB) memorandum to agencies on the use of AI also requires agencies to test their rights-impacting AI for significant disparities in its performance across demographic groups, and ensure the data used to test their AI is representative.[7] For example, the Centers for Medicare & Medicaid Services (CMS) would need to conduct a disparity assessment of an AI tool that, if adopted, significantly impacted Medicare eligibility decisions.

At the same time, data challenges have often made such assessments practically challenging. One interpretation of the data minimization principle — which holds that entities should collect only data minimally necessary to carry out a responsibility — has made equity assessments challenging for government agencies.[8] For example, the vast majority of agencies that published action plans

---

[6] "This protection should include proactive equity assessments as part of the system design, use of representative data and protection against proxies for demographic features, ensuring accessibility for people with disabilities in design and development, pre-deployment and ongoing disparity testing and mitigation, and clear organizational oversight. Independent evaluation and plain language reporting in the form of an algorithmic impact assessment, including disparity testing results and mitigation information, should be performed and made public whenever possible to confirm these protections." Off. Sci. & Tech. Policy, White House, Blueprint for an AI Bill of Rights (Oct. 4, 2022), https://www.whitehouse.gov/ostp/ai-bill-of-rights/. NIST AI Risk Management Framework recommends adopting practices to measure the impact of AI systems, including that "Appropriateness of AI metrics and effectiveness of existing controls are regularly assessed and updated, including reports of errors and potential impacts on affected communities." National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," January 2023, 29, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/LX34-FGZU]. EO 14,091 has a number of places in which it indicates the importance of equity assessments. E.g., sec. 1: "Executive departments and agencies (agencies) have engaged in historic work assessing how their policies and programs perpetuate barriers for underserved communities and developing strategies for removing those barriers" and sec. 10(e): "The term 'equitable data' refers to data that allow for rigorous assessment of the extent to which Government programs and policies yield consistently fair, just, and impartial treatment of all individuals."

[7] Agencies are required to take additional steps before initiating use of new or existing rights-impacting AI. These include: "1. Identify and document in their AI impact assessment when using data that contains information about a class protected by Federal nondiscrimination laws (e.g., race, age, etc.). Given the risks arising when AI may correlate demographic information with other types of information, agencies should also assess and document whether the AI model could foreseeably use other attributes as proxies for a protected characteristic and whether such use would significantly influence model performance; 2. Assess the AI in a real-world context to determine whether the AI model results in significant disparities in the model's performance (e.g., accuracy, precision, reliability in predicting outcomes) across demographic groups; 3. Mitigate disparities that lead to, or perpetuate, unlawful discrimination or harmful bias, or that decrease equity as a result of the government's use of the AI; and 4. Consistent with applicable law, cease use of the AI for agency decisionmaking if the agency is unable to adequately mitigate any associated risk of unlawful discrimination against protected classes. Agencies should maintain appropriate documentation to accompany this decision-making, and should disclose it publicly to the extent consistent with applicable law and governmentwide policy." Rights-impacting AI is defined as "AI whose output serves as a principal basis for a decision or action concerning a specific individual or entity that has a legal, material, binding, or similarly significant effect on that individual's or entity's: 1. Civil rights, civil liberties, or privacy, including but not limited to freedom of speech, voting, human autonomy, and protections from discrimination, excessive punishment, and unlawful surveillance; 2. Equal opportunities, including equitable access to education, housing, insurance, credit, employment, and other programs where civil rights and equal opportunity protections apply; or 3. Access to or the ability to apply for critical government resources or services, including healthcare, financial services, public housing, social services, transportation, and essential goods and services." OMB, M-24-10 Memorandum for the Heads of Executive Departments and Agencies on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, March 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.

[8] The Equitable Data Working Group, established pursuant to EO 13.895, noted that many agencies do not have access to data necessary for equity assessments, explaining that "[h]istorically, when specific

for advancing racial equity pursuant to EO 13,985 cited the lack of demographic data as a substantial barrier to conducting such assessments.[9]

Public Sector Data Challenges

The Privacy Act of 1974 and the Paperwork Reduction Act (PRA) of 1980 have limited government agency access to demographic data.[10] The Privacy Act of 1974 limits the federal government's ability to collect, use, or share — including to other government agencies — personally identifiable information (PII).[11] For example, the collection of demographic information has historically not been considered "necessary" for a government agency to administer a federal program (e.g., Medicare) and has thus been subject to collection, use, and sharing prohibitions under the Privacy Act.[12] The PRA makes the process of data collection methods difficult by requiring agencies to undergo lengthy review and approval processes.[13] To collect additional information, agencies must obtain both Privacy Act and PRA clearances.

The interaction of these two hurdles has meant that federal agencies often lack the data necessary to assess the impact that policies and programs have on different demographic groups. Until 2022, certain agencies used the questionable practice of having officials infer race and ethnicity by "visual observation" when not self-reported by participants in federal programs. For example, Department of Agriculture officials imputed race of participants in certain USDA programs[14] and Food and Nutrition Service officials visually assessed the race of individuals receiving Supplemental Nutrition Assistance Programs (SNAP).[15]

---

demographic data was not necessary for determining an individual's eligibility for a program, agencies did not collect such information." Equitable Data Working Group. 2002. "A Vision for Equitable Data Recommendations from the Equitable Data Working Group," 7, available at https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf.

[9] See King, Ho, Gupta, Wu & Webley-Brown, *supra* note 4 (explaining that of the 25 agencies that published equity action plans, 21 cited this challenge).

[10] See, id.; Equitable Data Working Group, *supra* note 8.

[11] "The Privacy Act of 1974 requires federal agencies to abide by a 'data minimization' principle, namely to: (a) collect personally identifiable information only as minimally necessary to carry out their statutory mission; (b) use the information only for its stated collection purpose; and (c) refrain from sharing or linking the data." Arushi Gupta, Victor Y. Wu, Helen Webley-Brown, Jennifer King, and Daniel E. Ho, The Privacy-Bias Tradeoff, Stanford Human-Centered AI Institute, Oct. 2023, 1, https://hai.stanford.edu/sites/default/files/2023-10/Privacy-Bias-Trade-Off.pdf. There are statutory exceptions to these disclosure limitations, including for enabling statistical research, benefitting an agency's mandate, or for "routine use" compatible with the purposes justifying the original data collection. For more information about the specific requirements of the Privacy Act, see, supra 6.

[12] See King, Ho, Gupta, Wu & Webley-Brown, *supra* note 4.

[13] "Under the Act, federal agencies adding new data collection efforts (such as surveys or web forms) are typically required to go through notice-and-comment and approval by the White House Office of Management and Budget (OMB)." Id. The Privacy Act provides mechanisms for enforcing proper use of federal data, including sanctions for violating approved uses. These include possible criminal penalties and fines for officials that improperly disclose or maintain records with personally identifiable information without prior notice. For an overview of criminal penalties under the Privacy Act, see https://www.justice.gov/opcl/overview-privacy-act-1974-2020-edition/criminal#:~:text=The%20Privacy%20Act%2 0allows%20for,if%20the%20official%20acts%20willfully.

[14] This policy applied to programs administered by the USDA as diverse as tenancy applications for multi-family housing projects to the Child and Adult Care Food Program.

[15] Id.; Food and Nutrition Service. 2022. Proposed Rule: SNAP-Revision of Civil Rights Data Collection Methods. https://www.fns.usda.gov/snap/fr-062722.

Recent important efforts have begun to address these challenges. As EO 13,985 acknowledges, "A first step to promoting equity in Government action is to gather the data necessary to inform that effort." An OMB study pursuant to EO 13,985 highlights the need for continued support and resources to address barriers to equity,[16] but more progress and specific guidance related to data collection approaches is needed given that the fair administration of federal programs and policies, and the realization of EO 13,895's legal obligations, requires the collection of demographic data necessary for disparity assessments. In general, agencies have adopted four approaches, each of which comes with limitations:

- **Surveys/Sampling:** A random subset of individuals may be selected to participate in a survey to provide race and ethnicity information. This method requires new data collection efforts, but reduces risks of misuse as the data is used for a targeted application. Response rates, however, can be quite low.

- **Form Collection:** Registration forms for government services may include demographic data fields that are either mandatory or voluntary to complete. While potentially more comprehensive, this method has the potential to suffer from nonresponse bias or reduce participation in government programs amongst eligible populations. Some research indicates respondents provide answers to certain sensitive questions at similar rates to other demographic data,[17] but other work suggests that non-response rates can be significant.[18]

- **Data linkage:** Agency data can be linked to existing demographic data. This method does not require new data collection, but requires technical infrastructure to operationalize and is subject to constraints under the Privacy Act.

- **Statistical Imputation:** Demographic information can be inferred using statistical methods or other information. This method uses existing data, avoiding costly data collection, but may introduce statistical biases.

---

[16] See Study to Identify Methods to Assess Equity: Report to the President, July 2021, https://www.whitehouse.gov/wp-content/uploads/2021/08/OMB-Report-on-E013985-Implementation_508-Complia nt-Secure-v1.1.pdf.

[17] Cahill, S., Singal, R., Grasso, C., King, D., Mayer, K., Baker, K., & Makadon, H. (2014). Do Ask, Do Tell: High Levels of Acceptability by Patients of Routine Collection of Sexual Orientation and Gender Identity Data in Four Diverse American Community Health Centers. *PloS one*, *9*(9), e107104. https://doi.org/10.1371/journal.pone.0107104. Fredriksen-Goldsen, K. I., & Kim, H.-J. (2015). Count Me In: Response to Sexual Orientation Measures Among Older Adults. *Research on Aging, 37*(5), 464–480. https://doi.org/10.1177/0164027514542109.

[18] Saunders, H. & Chidambaram, P. (2022). Medicaid Administrative Data: Challenges with Race, Ethnicity, and Other Demographic Variables. KFF. https://healthcare.rti.org/insights/improving-collection-of-self-reported-race-and-ethnicity-data.

Recent studies call attention to important gaps in both technical infrastructure and institutional resources to support agency data collection efforts.[19]

RECOMMENDATION

Federal agencies should proactively utilize, and leverage as appropriate, methods by which they can overcome data collection challenges, with due consideration of risks including threats to individual privacy and data misuse, to conduct racial and gender disparity assessments in government programs and services. To do so, federal agencies should weigh the systemic risks of broadening data collection for equity assessments against the potential value of programmatic equity assessments for government programs and services.

This recommendation focuses on methods to facilitate the collection of race, ethnicity, and gender data for the purposes of conducting disparity assessments, while maintaining the commitment to individual privacy under the Privacy Act. Although this recommendation focuses on race, ethnicity, and gender information, it can inform disparity assessments for other protected characteristics and demographic categories.

The practices below detail potential measures agencies can take to gather the information necessary to conduct gender and racial disparity assessments, while protecting individual privacy through institutional safeguards. This is not intended to provide an exhaustive list of approaches for collecting and safeguarding sensitive data, but instead provides a summary of specific risk-minimizing practices for agencies to explore, including to leverage where appropriate and to make changes necessary for their utilization. Risks to be weighed against benefits include threats to individual privacy and the potential for data misuse, particularly in relation to vulnerable groups.[20] These recommendations are primarily focused on the predominant tabular setting faced by federal agencies.

Overcoming Data Challenges. First, the Office of Management and Budget (OMB) and agencies should consider streamlining PRA approval for data collection for disparity assessments. The PRA exempts certain agency actions from its requirements and provides forms of clearance that agencies could utilize to collect race, ethnicity, and gender data for some set of individuals. For instance, PRA "generic clearance" allows agencies to receive approval from the Office of Information and Regulatory Affairs (OIRA) to conduct more than one information collection using similar methods, such as a survey. Although the initial collection must go through the normal PRA process, the subsequent and associated

[19] For example, the Veterans' Affairs administration relies on outdated and fragmented technical systems. For further discussion see King, Ho, Gupta, Wu & Webley-Brown, *supra* note 4.
[20] For example, linked data sets could allow for greater granularity in individual and group demographic identification.

collections require significantly less procedure to obtain approval.[21] Thus, the "generic clearance" can enable agencies to develop approaches for how best to assess disparities.[22] OMB and agencies should explore such options[23] for ensuring that the PRA is not a barrier to protecting civil rights.

Second, the OMB should consider how agency collection of basic demographic data can be consistent with the Privacy Act's data minimization principle[24] and recognized methods and standards for privacy protection.[25] Given aforementioned legal obligations to prevent algorithmic discrimination, such data provides

---

[21] See id., 2-3; Types of PRA Clearance, Digital.gov, https://pra.digital.gov/clearance-types/.

[22] For instance, the Department of Education obtained generic clearance for design-improvement work and could use pre-existing clearance to fast track delivery.

[23] Two such options are worth noting here. First, the PRA allows agencies to make "non substantive" changes to its information collections (e.g., updating a form to use a newer race reporting standard or changing the wording of an already-approved collection question to improve accuracy) without seeking public comment. OMB, Memorandum for the Heads of Executive Departments and Agencies and Independent Regulatory Agencies on Flexibility under the Paperwork Reduction Act for Compliance with Information Collection Requirements, July 22, 2016, 4-5, https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/pra_flexibilities_memo_7_22_16_finall.pdf. Second, agencies can also use "common form clearance" to streamline information collections and facilitate data linkage across agencies. A "host" agency obtains normal clearance to collect the information for use by the "host" agency and other agencies that seek the information for the same purpose. OMB, Memorandum for the Heads of Executive Departments and Agencies and Independent Regulatory Agencies on Flexibility under the Paperwork Reduction Act for Compliance with Information Collection Requirements, July 22, 2016, 4, https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/pra_flexibilities_memo_7_22_16_finall.pdf; Types of PRA Clearance, Digital.gov, https://pra.digital.gov/clearance-types/. Thus, agencies can create form memos (e.g., memos of approved use, memos of understanding) that enable information collected by one agency to be used by other agencies or government-wide.

[24] While data collection has the potential to harm vulnerable groups without safeguards or careful risk assessment (see William Seltzer and Margo Anderson. 2001. "The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses." *Social Research* 68(2): 481-513.), government statistics can hold tremendous benefits by exposing inequitable access to services and thus providing information critical for promoting equity. For a discussion of the benefits that statistical information can provide government entities and the policy making process, see Brian A. Harris-Kojetin and Constance F. Citro (Eds.). 2021. *Principles and Practices for a Federal Statistical Agency*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education; see also, Equitable Data Working Group, *supra* note 8.

[25] In addition to the recommendations herein, agencies must follow OMB's revisions to the Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15), March 28, 2024, https://www.whitehouse.gov/omb/briefing-room/2024/03/28/omb-publishes-revisions-to-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity/. OMB should also consider existing standards and methods employed by government agencies and private sector entities, such as the NIST Privacy Framework, ISO/IEC 29100:2011 – Information technology – Security techniques – Privacy framework, the ISO/IEC's information security standards and privacy information management extensions, NIST's Guide to Protecting the Confidentiality of PII. NIST. 2020. "The NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management." Version 1.0, https://www.nist.gov/privacy-framework/privacy-framework; ISO/IEC 29100:2011 – Information technology – Security techniques – Privacy framework, https://standards.iso.org/ittf/PubliclyAvailableStandards/c045123_ISO_IEC_29100_2011.zip; ISO/IEC 27001:2022 - Information security, cybersecurity and privacy protection (2022), https://www.iso.org/standard/27001; ISO/IEC 27701:2019 - Security techniques, Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management, https://www.iso.org/standard/71670.html; NIST. (2010). "Special Publication 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): Recommendations of the National Institute of Standards and Technology," https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf; HHS. (2012). "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," available at https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

significant opportunities for identifying and rectifying civil rights violations.[26] Specifically, the OMB should consider, consistent with Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity,[27] developing guidance on (1) circumstances under which the collection of basic demographic data is necessary to conduct equity assessments; and (2) the specific criteria, including the use of strategies that enable statistical inference[28], that permit the lawful collection of such data.[29]

Institutional Safeguards. Demographic information can, of course, pose some risks, and agencies[30] should develop improved practices to guard against misuse of demographic data. These could include:

- **Opting out:** When collecting information with administrative forms, agencies may consider allowing individuals to opt out of responding.

- **Disclose Privacy Protections:** Agencies may disclose the measures they have taken to safeguard sensitive data to foster trust and transparency.

- **Firewalls:** Agencies may consider creating a firewall between demographic data for disparity assessment and operational data. In other words, the office administering the program (e.g., making benefits decisions) can be prohibited from observing demographic data, which is used by a distinct office conducting the disparity assessment. This was the case, for instance, for the Treasury Department's assessment of disparities in tax administration, where demographic data was not shared with the office conducting audit selection in the Internal Revenue Service.[31] When using forms, agencies may wish to use

---

[26] For example, the Equitable Data Working Group, established by EO 13,895, noted recommended "[e]xpand[ing] protected access to data for equity assessment" as "[o]ne of the most common questions that agencies posed to the Working Group was how they could disaggregate their program participant rolls in order to identify and rectify any inequity." Equitable Data Working Group, supra note 8, at 7-8.
[27] See OMB, *supra* note 25.
[28] For a discussion of common drawbacks of statistical disclosure control and differential privacy, see Aleksandra Slavkovic and Jeremy Seeman. 2023. "Statistical Data Privacy: A Song of Privacy and Utility." *Annual Review of Statistics and Its Application* 10: 189-218.
[29] In assessing the risks of data collection, key factors for agencies to consider include risks of misuse or re-identification that would undermine safety and privacy. When considering the potential for misuse, agencies should also consider the potential for future misuse. Recent and historical precedent demonstrate repeatedly that data collected for one purpose can, over time and with different oversight regimes, be repurposed and misused in ways that raise new barriers to equity and privacy. See, e.g., "City's IDNYC Smart Card Chip Plan Slammed as Security Risk" https://www.thecity.nyc/2019/09/12/city-s-idnyc-smart-card-chip-plan-slammed-as-security-risk/. In addition, some categories of demographic data are more sensitive and pose a greater risk for re-identification. For example, studies have indicated that questions about citizenship status may decrease the response rate for a survey or form, given political discourse around immigration and fear of retaliation. Agencies should begin with demographic data that is coarse, due to lower risk of re-identification. Coarse race, ethnicity, and gender data is generally seen as less sensitive, relative to other categories. See, e.g., Hansi Lo Wang. 2018. Citizenship Question May Be 'Major Barrier' To 2020 Census Participation. NPR (Nov. 2018). https://www.npr.org/2018/11/01/663061835/citizenship-question-may-be-major-barrier-to-2020-census-participation; Matthew D Ingber, Andrew J Pincus, Michael B Kimberly, and Colleen M Campbell. [n. d.]. United States Department of Commerce vs. State of New York. https://www.supremecourt.gov/DocketPDF/18/18-966/95014/20190401170915326_18-966.bsac.pdf.
[30] Because of the focus on the Privacy Act and Paperwork Reduction Act, this recommendation focuses on the public sector, but these practices are also relevant for the private sector.
[31] *See* King, Ho, Gupta, Wu & Webley-Brown, *supra* note 4.

a separate form for demographic information to structurally ensure that such information will not be used for program administration.

- **Sandboxes**. In instances where administrative data is linked, agencies may consider sandbox environments through memorandums of understanding (MOUs) and inter-agency agreements for hosting and sharing data, with researchers conducting the disparity assessment being provided only information necessary to conduct the assessment (e.g., a form of least privilege access).. For instance, such an approach has been frequently used to link to Census or Bureau of Labor Statistics data.[32]

Agencies should also ensure the effectiveness and verifiability of the above mechanisms.

While these issues may be clearest in the context of federal public sector data collection efforts due to experience with the Privacy Act and PRA, similar challenges and tradeoffs exist for state and local government agencies[33] and in the private sector. For instance, voluntary commitments made by companies to promote safe, secure, and transparent development and use of AI include commitments to assess the bias of AI systems.[34] New York City recently enacted legislation requiring private employers using commercial automated decision systems for hiring or promotion to conduct annual independent audits that assess fairness of these systems across race and gender.[35] The practices detailed are thus also relevant for private sector entities trying to navigate data collection necessary to assess algorithmic discrimination, while ensuring adequate privacy protections.

---

[32] For example, student debt relief policies involved user testing and service delivery, which involved back-end coordination between the Department of Education and the Internal Revenue Service for data matching purposes. Computer Matching Agreement Between U.S. Department of Education (ED) and U.S. Department of the Treasury Internal Revenue Service (IRS) Future Act Direct Data Exchange (FA-DDX), 2023, https://www2.ed.gov/about/offices/list/om/docs/pirms/future-cma.pdf. See also, Section VI. Legal Authorization in Guidance for Providing and Using Administrative Data for Statistical Purposes, OMB, February 14, 2014, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

[33] The Privacy Act and PRA apply to federal government agencies not state and local government agencies.

[34] Voluntary commitments secured by the Biden-Harris administration in July 2023 include several provisions that implicate data collection efforts. For instance,"prioritizing research on societal risks posed by AI systems, including on avoiding harmful bias and discrimination, and protecting privacy." See Ensuring Safe, Secure, and Trustworthy AI, White House, July 21, 2023, https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

[35] NYC Local Law 144 requires bias audits of automated employment decision tools used by employers to be publicly disclosed. See Jacob Metcalf. 2023. What federal agencies can learn from New York City's AI hiring law. The Hill (Dec. 2023) https://thehill.com/opinion/technology/4360523-what-federal-agencies-can-learn-from-new-york-citys-ai-hiring-law /.

ACKNOWLEDGEMENTS

The NAIAC Safety, Trust, and Rights working group participated in the preparation of this document. Contributors include:

- Paula Goldman
- Janet Haven
- Daniel E. Ho
- Ashley Llorens
- Christina Montgomery
- Liz O'Sullivan

The working group thanks many individuals who joined us for briefings and Christie Lawrence and Lindsey Gailmard for help in developing this recommendation. A quorum of the membership of NAIAC reviewed and approved this document.

ABOUT NAIAC

The National Artificial Intelligence Advisory Committee (NAIAC) advises the President and the White House National AI Initiative Office (NAIIO) on the intersection of AI and innovation, competition, societal issues, the economy, law, international relations, and other areas that can and will be impacted by AI in the near and long term. Their work guides the U.S. government in leveraging AI in a uniquely American way — one that prioritizes democratic values and civil liberties, while also increasing opportunity.

NAIAC was established in April 2022 by the William M. (Mac) Thornberry National Defense Authorization Act. It first convened in May 2022. It consists of leading experts in AI across a wide range of domains, from industry to academia to civil society. https://www.ai.gov/naiac/

###