Stanford University
Human-Centered
Artificial Intelligence

Policy Brief
HAI Policy & Society
October 2023

The Privacy-Bias Trade-Off

Arushi Gupta, Victor Y. Wu, Helen Webley-Brown, Jennifer King, and Daniel E. Ho

ALGORITHMIC FAIRNESS AND PRIVACY ISSUES are increasingly drawing both policymakers' and the public's attention amid rapid advances in artificial intelligence (AI). New AI applications used in medicine, criminal justice, hiring, and elsewhere can—and in numerous cases already do—make decisions that can generate or exacerbate disparities along race or gender attributes. At the same time, the vast amounts of data collected and processed by public and private actors to train models carry complex implications for individual privacy.

Safeguarding privacy and addressing algorithmic bias can pose a less recognized trade-off. The principle of "data minimization," which the U.S. government has experimented with for almost 50 years, holds that entities should collect and retain only the minimally necessary data to achieve their objectives. But the result is that agencies may lack access to demographic data (e.g., data on race, ethnicity) that is required to conduct equity assessments of public programs. Privacy, in short, can mean a lack of awareness.

In a new paper, "The Privacy-Bias Trade-Off," we document this tension between data minimization principles and racial disparity assessments in the U.S. government. We examine the U.S. government's recent efforts to introduce government-wide equity assessments of federal programs and consider a range of policy solutions, including amending or interpreting the Privacy Act to permit the collection of demographic data to conduct disparity assessments.

Key Takeaways

As companies and regulators step up efforts to protect individuals' information privacy, a common privacy principle (data minimization) can come to clash with algorithmic fairness.

The U.S. federal government provides a compelling case study: Its adoption of data minimization in the Privacy Act of 1974 has brought many privacy benefits but stymies efforts to gather demographic data to assess disparities in program outcomes across federal agencies.

Coupled with procedures under the Paperwork Reduction Act of 1980, the Privacy Act has meant that agencies rarely and inconsistently collect data on protected attributes.

Twenty-one of 25 agencies noted substantial data challenges in responding to an executive order requiring agencies to conduct equity assessments of their programs.

Privacy principles should be harmonized to permit secure collection of demographic data to conduct disparity assessments.

Data minimization, while beneficial for privacy, has simultaneously made it legally, technically, and bureaucratically difficult to acquire demographic information necessary to conduct equity assessments. The more AI systems are deployed across government and society, the more imperative it will be to balance privacy and fairness.

Introduction

Race and ethnicity are socially constructed concepts, but demographic information remains critical to understanding, let alone mitigating, racial (or intersectional) disparities. As put elegantly in the algorithmic fairness literature, there is no fairness without awareness.

Approaches to measuring race and ethnicity have varied over time. The U.S. government, for example, started collecting race-related data in the 1930s from applicants to the Social Security Administration (SSA); prior to 1980, the categories were “White,” “Negro,” and “Other.” Since then, the SSA has repeatedly changed its race/ethnicity codes. Most recently, the country’s chief statistician announced plans in 2022 to update government guidelines to further disaggregate racial categories; these include moving away from monolithic categories such as “Asian American,” which can mask substantial variance between subgroups.

During his first day in office, President Biden signed Executive Order (EO) 13985, requiring agencies to conduct equity assessments of federal policies and programs. The EO also acknowledged that “[m]any Federal datasets are not disaggregated by race, ethnicity, gender, disability, income, veteran status, or other key demographic variables.”

The U.S. government’s implementation of data minimization is a potent case study that highlights a broad dilemma that has vexed regulators and industry alike: Data minimization can come to clash with disparity or equity assessments.

What it did not mention is one key structural reason for such difficulties: privacy protection.

The Privacy Act of 1974 and the Paperwork Reduction Act of 1980 are laws that aim to reduce the amount of data the government collects. 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. The Paperwork Reduction Act of 1980 requires agencies to secure approval before requesting many kinds of data from the public. Under the Act, federal agencies adding new data collection mechanisms (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).

The reality is very different: Demographic data collection by government agencies about race and ethnicity remains inconsistent and often of poor quality.

The U.S. government's implementation of data minimization is a potent case study that highlights a broad dilemma that has vexed regulators and industry alike: Data minimization can come to clash with disparity or equity assessments. Existing research has examined tensions between privacy and fairness in the algorithmic context. It has also examined how privacy laws, such as the European Union's General Data Protection Regulation (GDPR), can inadvertently prevent technologists from accessing the data they need to conduct fairness tests. Missing, though, is research that documents the impact of privacy-fairness trade-offs on government policy and data use, where the widely espoused data minimization principle has been adopted for some 50 years.

In our paper, we examine how agencies are grappling with this "privacy-bias trade-off." We assess agency responses to the EO's requirement to conduct equity assessments and how agencies have dealt with privacy and fairness considerations in major claims programs. We outline federal agencies' distinct approaches to data surveying and management and identify the most common barriers to implementing equity assessments.

Research Outcomes

On the surface, there is widespread support from policymakers for conducting disparity assessments, as evident from provisions in the Affordable Care Act and the Consumer Financial Protection Bureau's 2017 policy change to permit more demographic data collection. But the reality is very different: Demographic data collection by government agencies about race and ethnicity remains inconsistent and often of poor quality.

Of the 25 agencies filing equity action plans in response to the racial justice EO, 21 noted the lack of demographic data as a substantial barrier. Only some of the plans that mention demographic data provide concrete solutions: Two agencies aim to run surveys, eight plan to change public-facing forms, four propose record linkage, and one plans to use imputation to tackle this issue. Thirteen agencies, though, shared only partial or generic descriptions of planned changes.

Prior attempts to conduct similar equity assessments illustrate the many barriers emanating from the privacy-bias trade-off. First, and most important, some laws directly limit the collection of demographic data. The Equal Credit Opportunity Act, for instance, prohibits data collection on race and ethnicity for many agricultural loans and small business loans. The U.S. Treasury Department views the Internal Revenue Service as statutorily prohibited from linking census records for demographic attributes. And the Paperwork Reduction Act makes collection of demographic information by revising forms a prolonged process. The truly perverse result of such constraints is that some agencies resorted to inferring race by "visual observation." As late as 2022, for instance, the U.S. Department of Agriculture imputed race by "visual observation" when race information was not collected.

Of the 25 agencies filing equity action plans in response to the racial justice EO, 21 noted the lack of demographic data as a substantial barrier.

Second, data minimization has meant that agency datasets exist in silos. Fragmented and outdated technical infrastructure and lack of technical expertise for how to update such systems have impeded disparity assessments. For instance, a 2011 Department of Agriculture regulation prohibited employees from visually observing race and ethnicity, such as when entering loan data, but its customer data management system required a race data field as late as 2019. It took the agency months to fully update its data management system to fix the issue. The Department of Veterans Affairs, meanwhile, collects demographic data but does not have the data synchronized in a centralized system, instead relying on what researchers have called a "sometimes-confusing alphabet soup of data partners."

Third, we document considerable resistance to data collection among federal agencies or private third-party data collectors due to public relations, political, or litigation risks. The Privacy Act of 1974, for instance, does not explicitly list bias as one of its statutory exceptions to data disclosure limits. Bias does not fit easily into any of the three exempt categories, which include enabling statistical research, benefiting an agency's mandate (the Act has a "need to know" provision), and "routine use" otherwise incompatible with the purpose of the data's initial collection. In fact, when the U.S. Commission on Evidence-Based Policymaking issued a 2018 survey of federal agencies, 47 of 79 respondents across agencies said data linkage was constrained by "statutes prohibiting data sharing"; 66 respondents said it was other "regulations and policies that make it difficult to link data." Several other laws, such as the E-Government Act of 2002, place limits on data collection and sharing.

Last, stakeholders express genuine concern that asking for demographic data will increase survey non-response rates or affect program participation. They worry that mandatory and even voluntary data collection pose the risk of making respondents uncomfortable, lowering survey response rates, and potentially engendering distrust of how such information would be used.

Policy Discussion

So how can the federal government resolve this privacy-bias trade-off? We make three proposals that preserve the Privacy Act's core principles but would enable agencies to conduct meaningful equity assessments.

First, the Privacy Act should be interpreted or amended to permit interagency record linkage for bias assessments. One rationale is that demographic attribute information is minimally required for agencies to provide their services equitably, and hence meets the data minimization principle. Another is that the Privacy Act's "statistical research" and "routine use" exceptions can be interpreted to include bias assessments.

Second, agencies should adopt strict institutional protections so that demographic data is used exclusively for equity assessments. Internally separating investigatory and adjudicatory functions is a long-standing feature of administrative law; a similar principle should be promulgated so that a unit conducting an equity assessment within an agency is institutionally separated from the unit administering the program. Census data, for instance, could be provided only to the Treasury Department's Office of Tax Analysis and not the Internal Revenue Service's audit teams. Such protections would promote public trust, protect privacy, and address the most significant concerns about the abuse of demographic data.

Third, the Paperwork Reduction Act should be interpreted or amended to streamline the process of gathering demographic data on forms and surveys. When OMB has already promulgated a race reporting standard, there is little benefit from the procedural machinations that lead to months of delay to collect such information.

Fourth, Congress should enact and scale initiatives like the National Secure Data Service and the National AI Research Resource that advance privacy-protective sharing of administrative data. Such initiatives would address the data fragmentation and technical limitations that have impeded equity assessments.

The privacy-bias trade-off has a profound impact on agencies' ability to conduct equity assessments, but these simple reforms would enable agencies to navigate that trade-off rather than turning a blind eye to disparities in the name of privacy.

We close by noting that the privacy-bias trade-off is not unique to government agencies. The public sector simply provides the best evidence from its 50-year experiment in data minimization. Similar issues will present themselves in private sector data, where responsible resolution of the trade-off will be an increasingly important policy issue.

> *The privacy-bias trade-off has a profound impact on agencies' ability to conduct equity assessments, but these simple reforms enable agencies to navigate that trade-off rather than turning a blind eye to disparities in the name of privacy.*

Reference: The original article is accessible at Arushi Gupta et al., "**The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government**," *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (June 2023): 492-505, https://dl.acm.org/doi/10.1145/3593013.3594015. The article received a Best Paper Award at FAccT '23.

Stanford University's Institute on Human-Centered Artificial Intelligence (HAI) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu.**

Arushi Gupta graduated from Stanford University with a B.A. in political science and an M.S. in computer science.

Victor Y. Wu is a J.D. candidate at Stanford Law School.

Helen Webley-Brown is a Ph.D. student in American politics and political methodology at MIT.

Jennifer King is the Privacy and Data Policy Fellow at the Stanford Institute for Human-Centered Artificial Intelligence.

Daniel E. Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, Professor of Computer Science (by courtesy), Senior Fellow at the Stanford Institute for Economic Policy Research, and Director of the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University.

Stanford University
Human-Centered
Artificial Intelligence

Stanford HAI: 353 Jane Stanford Way, Stanford CA 94305-5008
T 650.725.4537 **F** 650.123.4567 **E** HAI-Policy@stanford.edu **hai.stanford.edu**