



December 5, 2023

Office of Management and Budget
725 17th Street NW
Washington DC, 20503

Submitted via regulations.gov

Re: OMB-2023-0020 — Request for Comments on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence Draft Memorandum.

We write to provide comments in response to the Office of Management and Budget’s draft memorandum, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (AI)*.

Upturn is a non-profit organization that advances equity and justice in the design, governance, and use of technology. Through research and advocacy, we drive policy change by investigating specific ways that technology and automation shape people’s opportunities, particularly in historically disadvantaged communities.

Our comments primarily address questions 5, 6, 7, and 8 in the request for comment.

Executive Summary

The Office of Management and Budget’s draft memorandum, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (AI)*, has the potential to help prevent and address discrimination in the use of automated systems by federal agencies. By requiring anti-discrimination testing of a broad range of rights-impacting algorithmic systems, as well as ongoing monitoring and mitigation of algorithmic discrimination, the memorandum will launch a landmark effort to evaluate algorithmic systems in civil rights areas — a framework that Upturn has advocated for in many civil rights contexts such as credit, employment, housing, and policing. This effort can materially improve peoples’ lives, especially for marginalized communities protected by federal anti-discrimination laws. As one example, algorithmic testing has identified methods to mitigate pronounced racial disparities in IRS models used to select individuals for tax audits. By committing agencies to perform anti-discrimination testing of their algorithmic systems, the federal government can “serve as a model for state and local governments, businesses and others to follow in their own procurement and use of AI.”¹ The final memorandum must require agencies to perform anti-discrimination testing of their systems and mitigate disparate impact.

However, these important measures risk being undercut by other provisions of the draft memorandum. In particular, the draft memorandum affords agencies significant leeway to waive compliance with the minimum practices. The Office of Management and Budget (OMB) should ensure that agencies, unless expressly and strictly prohibited by statute, explore ways to safely collect or infer the necessary demographic data to comply with the memorandum’s minimum requirements. An agency should only be able to waive compliance with the memorandum’s anti-discrimination testing provisions if two conditions are met: first, an agency determines that a specific legal barrier prevents them from collecting relevant demographic data, and second, an agency makes a written determination that no other method to perform the anti-discrimination testing is viable. In a large majority of cases, other methods — beyond direct collection of self-reported demographic data — should be available to support these efforts. As a result, it should be the rare case that agencies are able to waive compliance with the memorandum’s anti-discrimination testing provisions.

¹ The White House, “FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans’ Rights and Safety,” May 4, 2023, available at <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administratio-n-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/>.

- 1. The final memorandum must contain two key provisions. First, agencies must perform anti-discrimination testing of their algorithmic systems. Second, agencies must be required to explore mechanisms to mitigate disparate impact.**

We applaud OMB’s draft memorandum for broadly defining “rights-impacting” algorithmic systems and requiring agencies to conduct anti-discrimination testing of these systems. We are also heartened to see that the draft memorandum would further require agencies to mitigate a system’s disparate impact, consistent with applicable law, once that disparate impact has been identified. It is critical that the provisions in Section 5(c)(v)(A)-(C) remain in OMB’s final memorandum. When the federal government uses algorithmic systems in covered civil rights areas, it must ensure that those systems are regularly tested for disparate effects on a prohibited basis. Similarly, agencies must maintain reasonable measures to search for less discriminatory algorithms on an ongoing basis. These provisions are consistent with the administration’s policy, as expressed through Executive Orders 14091 and 14110, as well as the AI Bill of Rights. Executive Order 14091 broadly required agencies to consider opportunities to “prevent and remedy discrimination, including by protecting the public from algorithmic discrimination.”² Executive Order 14110 stated the administration’s policy that 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,”³ and it more broadly directed agencies to use their authorities to prevent and address discrimination in the use of automated systems.⁴ The AI Bill of Rights called for designers, developers, and deployers of automated systems to “take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way,” and for “proactive equity assessments as part

² Executive Order 14091, Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government, 88 Fed. Reg. 10825, 10831, February 22, 2023, available at <https://www.federalregister.gov/documents/2023/02/22/2023-03779/further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal>.

³ Executive Order 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75191, 75192, November 1, 2023, available at <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>

⁴ See, e.g., Sections 6, 7, and 8 of Executive Order 14110.

of the system design,” as well as “pre-deployment and ongoing disparity testing and mitigation.”⁵

Such requirements are consistent with recent work by Upturn and our co-authors that argues that the duty to search for less discriminatory algorithms should be on the entities that develop and deploy predictive models.⁶ In this case, that duty would fall to federal agencies and their contractors. An often unspoken premise throughout many efforts to regulate algorithmic systems is that for any given prediction problem, a single “correct” model exists. For example, when a bank seeks to predict default by borrowers, it is often assumed that a single “correct” model exists that best advances that goal, and that any deviation from this unique solution would necessarily entail a loss in performance. The implication is that pursuing goals like minimizing disparate impact will inevitably involve a tradeoff with model performance. But the assumption that a unique solution exists and that a fairness-accuracy tradeoff is inevitable are descriptively inaccurate. Work in computer science has established that there are almost always multiple possible models with equivalent accuracy for a given prediction problem — a phenomenon termed “model multiplicity.”⁷

Multiplicitous models perform a given prediction task equally well, but can differ in other ways — from the features they use to make predictions, to the way they combine those features to make predictions, to the way their predictions are robust to changing circumstances. Critically, these equally performant models can have different levels of disparate impact. As a result, when an algorithmic system displays a disparate impact, model multiplicity suggests that other models that perform equally well, but have less

⁵ White House Office of Science and Technology Policy, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (Oct. 2022), available at <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>.

⁶ See Emily Black, John Logan Koepke, Pauline T. Kim, Solon Barocas, Mingwei Hsu, *Less Discriminatory Algorithms*, Oct. 2, 2023, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590481.

⁷ Several different terms have been used to describe related phenomena over years of computer science and statistical scholarship. The first to introduce the notion that various models could be equally effective at the same task was Leo Breiman. See *Statistical Modeling: The Two Cultures*, 16 *Stat. Sci.*, no. 3, Aug. 2001 at 199, 200 (using the term “the Rashomon effect”). After this, Marx et al. resurfaced the idea that different models could have different predictions but similar performance, under the term “predictive multiplicity.” See Charles Marx et al., *Predictive Multiplicity in Classification*, 119 *Proc. Machine Learning Research*. 6765 (2020). Black and Fredrikson displayed similar behavior on different classes of models in concurrent work. See Emily Black, Matt Fredrikson, *Leave-One-Out Unfairness*, in *FACCT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 285 (2021). Later, Black et al., introduced the term model multiplicity to encompass not only how similarly performant models differ in their predictions, but also in their internals, which have impacts on the explanations of their predictions. See Emily Black, Manish Raghavan, Solon Barocas, *Model Multiplicity: Opportunities, Concerns, and Solutions*, in *FACCT ’22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* 850 (2022).

discriminatory effect, exist. In other words, in almost all cases, a less discriminatory algorithm (LDA) exists.

These insights about model multiplicity have profound ramifications for the legal, regulatory, and policy response to discriminatory algorithms and support OMB’s anti-discrimination testing provisions. Under disparate impact doctrine, it makes little sense to say that a given algorithmic system is either “justified” or “necessary” if an equally accurate model that exhibits less disparate effects is available and discoverable with reasonable efforts. In fact, a close reading of the legal authorities over the decades reveals that the law has on numerous occasions recognized that the existence of a less discriminatory alternative is sometimes relevant to a defendant’s burden of justification at the second step of disparate impact analysis.⁸ As a result, when entities, including the federal government, use algorithmic systems in civil rights domains, they should have a duty to search for and implement LDAs before they can deploy a system with disparate effects. Without such a duty, developers are likely to be singularly focused on their chosen performance metric and will fail to identify ways to achieve the same goals with less discriminatory impact. OMB’s memorandum is on solid legal and technical footing when it places this duty on federal agencies and contractors who develop and deploy rights-impacting algorithmic systems.

Imposing such a duty not only comports with the purposes behind our civil rights laws, which are intended to remove arbitrary barriers to full participation by marginalized groups in our nation’s economic life, but also is practical, because model developers are in the best position to undertake a fruitful search for LDAs. Developing a model through the machine learning pipeline inherently involves testing and exploration of alternatives. A requirement that entities, such as federal agencies or their contractors, also test for disparate impact and compare model disparities throughout the model development process is straightforward and is not, by itself, burdensome.

Notably, this approach differs from past attempts to combat disparate impact, which would have required entities to prove the absence of less discriminatory alternatives in justifying their challenged practice. Historically, such approaches were

⁸ Emily Black, John Logan Koepke, Pauline T. Kim, Solon Barocas, Mingwei Hsu, *Less Discriminatory Algorithms*, Oct. 2, 2023, at 16-28.

critiqued for requiring entities to prove a negative.⁹ But a requirement that entities, including federal agencies, maintain reasonable steps to search for and implement LDAs is different. For one, there is functionally no uncertainty as to whether an LDA exists and there is a structured process for discovering them. For another, there are methods to quantify model properties, such as model performance, so as to make the baseline and alternative directly comparable. Moreover, it is unlikely that a developer has, without any specific exploration or dedicated process, randomly happened upon the globally optimal, least discriminatory model. As a result, OMB is justified in requiring federal agencies and their contractors to test their models for disparate impact and search for ways to mitigate disparate impact if it is identified.

2. For agencies to fulfill the “Additional Minimum Practices for Rights-Impacting AI” in 5(c)(v), they will need to meet certain basic requirements.

As currently written, the draft memorandum would require agencies to abide by a number of minimum practices for rights-impacting AI. For example, once designated rights-impacting, agencies will need to “assess whether their rights-impacting AI materially relies on information about a class protected by Federal nondiscrimination laws in a way that could result in algorithmic discrimination or bias against that protected class,”¹⁰ “test their AI to determine whether there are significant disparities in the AI’s performance across demographic groups,”¹¹ and “appropriately address disparities that have the potential to lead to discrimination, cause meaningful harm, or decrease equity, dignity, or fairness.”¹² The draft memorandum also calls on agencies to stop using rights-impacting AI systems if “adequate mitigation of the disparity is not possible.”¹³

⁹ For example, when the Department of Housing and Urban Development reinstated their 2013 “Implementation of the Fair Housing act’s Discriminatory Effects Standard” in 2023, the Department noted that their approach — adopting Title VII’s burden-shifting framework — “makes the most sense because it does not require either party to prove a negative.” See Department of Housing and Urban Development, “Reinstatement of HUD’s Discriminatory Effects Standard,” 88 Fed. Reg. 19450, 19490, March 31, 2023, available at <https://www.federalregister.gov/documents/2023/03/31/2023-05836/reinstatement-of-huds-discriminatory-effects-standard>.

¹⁰ Executive Office of the President, Office of Management and Budget, Proposed Memorandum for the Heads of Executive Departments and Agencies, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, Section 5(c)(v)(A)(1) at 18, available at <https://ai.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-Public-Comment.pdf>.

¹¹ Id., Section 5(c)(v)(A)(2) at 18.

¹² Id.

¹³ Id.

For agencies to fulfill these minimum practices, they will need to ensure that the following four related processes are in place.

1. Agencies must have a process in place to collect or infer the demographic data necessary to perform a disparate impact analysis. For example, absent information about the gender of people whose data is being used to evaluate a model's performance, developers will be unable to establish whether the model's performance and selection rate differs by gender.
2. Agencies must have a process in place for actually performing a disparate impact analysis. Notably, this must include a process for evaluating a model for disparate impact both prior to deployment and on an ongoing basis, once it has been deployed.¹⁴
3. Agencies must establish a process for searching for LDAs. This should apply to models being developed for the first time — where the search for LDAs can be incorporated into the model development process from the outset — and in addressing a disparate impact that has been identified after a model has been developed or deployed.
4. Agencies must establish processes to determine when they will adopt an LDA and for implementing the LDA in practice.

Absent any one of these processes, agencies will fail to fulfill the minimum requirements. In the final memorandum, or through other guidance to agencies, OMB should consider clarifying that it expects each of these four related processes to be in place for agencies to fulfill the minimum practices.

To ensure that agencies are best able to advance anti-discrimination testing of algorithmic systems, OMB should clarify that agencies, unless expressly and strictly prohibited by statute, should explore ways to safely collect or infer the necessary demographic data to comply with the memorandum's minimum requirements. In particular, some agencies may believe that they cannot effectively comply with the minimum requirements because they do not currently collect or infer the relevant demographic data necessary to perform anti-discrimination testing.¹⁵ Agencies may point to a variety of reasons why they currently do not collect or infer relevant demographic information: a relevant statute may clearly prohibit the agency from directly collecting

¹⁴ Section 5(c)(v)(C) at 20.

¹⁵ See, e.g., Arushi Gupta, Victor Y. Wu, Helen Webley-Brown, Jennifer King, Daniel E. Ho, *The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government*, in FAcCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency 492 (2023).

demographic data, a statute may prohibit agencies from sharing relevant information, agencies may have an institutional norm against collecting demographic data, or agencies may have limited experience in applying relevant inference methodologies. OMB should clarify that it expects agencies, where permissible under existing law, to make every effort to re-examine agency-level policies, directives, regulations, practices, or norms that would hinder them from performing anti-discrimination testing of their algorithmic systems. Such efforts are directly responsive to Executive Orders 14091 and 14110, the AI Bill of Rights, and the recommendations from the Equitable Data Working Group. And a number of agencies have experience and practice in inferring demographic data for anti-discrimination testing purposes when that data cannot be directly collected.¹⁶

One reason that agencies should be expected to make every effort to re-examine existing policies, regulations, directives, practices, or norms that would hinder anti-discrimination testing is that, currently, the draft memorandum states that “[e]xcept as prevented by applicable law and governmentwide guidance, agencies must apply the

¹⁶ For example, the CFPB’s Office of Research and Division of Supervision, Enforcement, and Fair Lending rely “on a BISG proxy probability for race and ethnicity in fair lending analysis conducted for non-mortgage products.” See Consumer Financial Protection Bureau, *Using publicly available information to proxy for unidentified race and ethnicity: A methodology and assessment*, at 23 (2014), available at https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf. The CFPB also has a gender proxy methodology. See Consumer Financial Protection Bureau, *Semi-Annual Report of the Consumer Financial Protection Bureau*, Fall 2021, at 39, available at https://files.consumerfinance.gov/f/documents/cfpb_semi-annual-report-to-congress_2022-04.pdf. The FTC also relies on BISG/BIFSG in their research efforts. See Federal Trade Commission, *Protecting Older Consumers 2021-2022: A Report of the Federal Trade Commission*, at 39 (Oct. 18, 2022), available at https://www.ftc.gov/system/les/ftc_gov/pdf/P144400OlderConsumersReportFY22.pdf. The Centers for Medicare and Medicaid Services uses a specialized version of BISG, Medicare BISG, or MBISG. MBISG is currently used to conduct national, contract-level, stratified reporting of Medicare Part C & D performance data for Medicare Advantage Plans by race and ethnicity. See Centers for Medicare and Medicaid Services, *The Path Forward: Improving Data to Advance Health Equity Solutions* (November 2022), available at <https://www.cms.gov/files/document/path-forwardhe-data-paper.pdf>. The EEOC’s Investigative Analytics Team uses BISG race estimation techniques when race/ethnicity is missing from administrative employment data provided by employers. See “Using Bayesian Improved Surname Geocoding (BISG) to Classify Race and Ethnicity in Administrative Employment Data by Industry: A Validation Study,” available at <https://www2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=311006>. The Office of the Assistant Secretary for Planning and Evaluation and the Centers for Medicare and Medicaid Services have funded research supporting the development and advancement of methods like BIFSG. See Melony E. Sorbero, Roald Euler, Aaron Kofner, Marc N. Elliott, *Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015-2022 Open Enrollment Periods*, (2022) available at https://www.rand.org/content/dam/rand/pubs/research_reports/RR1800/RR1853-1/RAND_RR1853-1.pdf. And the Department of the Treasury recently relied upon BIFSG for the first time to conduct tax analysis by race and Hispanic origin. See Robin Fisher, *Estimation of Race and Ethnicity by Re-Weighting Tax Data*, Department of the Treasury, Office of Tax Analysis, Technical Paper 11, January 2023, available at <https://home.treasury.gov/system/files/131/TP-11.pdf>; also see Julie-Anne Cronin, Portia DeFilippes, Robin Fisher, *Tax Expenditures by Race and Hispanic Ethnicity: An Application of the U.S. Treasury Department’s Race and Hispanic Ethnicity Imputation*, Department of the Treasury, Office of Tax Analysis, Working Paper 122, January 2023, available at <https://home.treasury.gov/system/files/131/WP-122.pdf>.

minimum practices in this section to safety-impacting and rights-impacting AI by August 1, 2024, or else stop using the AI until it becomes compliant.”¹⁷ As drafted, we take this provision to mean that if any agency claims that an existing statute prevents it from complying with the minimum practices, they do not necessarily have to stop using the AI system, even as it remains non-compliant. OMB should require agencies to specify exactly which provision of applicable law prevents them from applying the minimum practices. For example, if an agency determines that an existing legal barrier would prevent them from collecting the relevant demographic information to perform anti-discrimination testing of algorithmic systems, and separately also determines that no viable alternative methods to perform the testing are viable, the agency should be required to provide that determination to OMB in writing. The determination should also clearly state why other methods are insufficient to enable anti-discrimination testing.¹⁸ Absent specific legal prohibition or other governmentwide guidance, if an agency is unable to perform anti-discrimination testing of an algorithmic system, the agency must cease use of that system.

It is key that OMB not only require agencies to point to the legal barrier, but to also provide a detailed justification as to why no other viable alternative method would enable them perform the relevant anti-discrimination testing. Recent work on models used by the IRS to select individuals for audits provides a clear example of how agencies can perform anti-discrimination testing in the absence of directly collected demographic data.¹⁹ The goal of these models was to predict when an individual was at high risk of tax noncompliance. Because the IRS “does not systematically collect data on taxpayer race, either directly via tax returns or indirectly via merging tax data with administrative data on race from other agencies,” researchers turned to Bayesian Improved First Name Surname Geocoding (BIFSG) “to estimate the probability that a taxpayer is Black (and non-Hispanic) based on the first name, last name, and location of the taxpayer.”²⁰

As the researchers show, different problem formulations — the translation of a

¹⁷ Proposed Memorandum for the Heads of Executive Departments and Agencies, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, Section 5(c) at 13.

¹⁸ Especially when agencies seek to waive compliance with the minimum practices articulated in Section 5(c)(v) of the memorandum, that waiver must be public. This should include the agency’s rationale and justification for the waiver, including the specific provision of law that would prevent an agency from complying.

¹⁹ Emily Black, Hadi Elzayn, Alexandra Chouldechova, Jacob Goldin, Daniel E. Ho, *Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models*, in FAcCT ‘22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency 1479 (2022); Hadi Elzayn, Evelyn Smith, Thomas Hertz, Arun Ramesh, Robin Fisher, Daniel E. Ho, Jacob Goldin, *Measuring and Mitigating Racial Disparities in Tax Audits*, Stanford Institute for Policy Research Working Paper (2023), available at <https://siepr.stanford.edu/publications/working-paper/measuring-and-mitigating-racial-disparities-tax-audits>.

²⁰ Elzayn, et al., *Measuring and Mitigating Racial Disparities in Tax Audits*, at 19.

real-world problem into a machine learning task — can lead to different results. When the problem was formulated to predict whether individuals are likely to be noncompliant *at all* (with binary labels, describing if an individual was compliant or not) — as opposed to predicting the amount of money they failed to report (with continuous labels of the amount of taxes owed) — disproportionately more lower-income and Black individuals were selected for audit.²¹ As a result, changing the model’s prediction task from the *likelihood of noncompliance* to the *expected amount of noncompliance* shifted the distribution of those recommended for audit by the algorithm from lower-income and Black individuals towards higher-income and more white individuals, reducing stark disparities.²² Without BIFSG, the researchers would not have been able to perform the basic disparate impact testing, let alone search for an alternative approach that reduced disparities.

3. The final memorandum should ensure that Chief AI Officers do not have such wide latitude to invoke a waiver from the minimum practices for rights-impacting AI.

As currently drafted, OMB’s memorandum allows CAIOs to:

waive one or more of the [minimum practices for safety-impacting and rights-impacting artificial intelligence] for a specific covered AI application or component after making a written determination, based upon a system-specific risk assessment, that fulfilling the requirement would increase risks to safety or rights overall or would create an unacceptable impediment to critical agency operations.²³

The draft memorandum defines “waiving individual applications of AI from elements of Section 5 of this memorandum” as one of the responsibilities of a CAIO.²⁴ As drafted, these provisions would likely allow many safety- and rights-impacting algorithmic systems to evade scrutiny. OMB should make several changes to the memorandum to ensure that CAIOs do not routinely seek waivers and undermine the purpose of the memorandum.

²¹ Id., 36-37.

²² Id.

²³ Proposed Memorandum for the Heads of Executive Departments and Agencies, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, Section 5(c)(iii) at 14.

²⁴ Id., Section 3(b)(ii)(O) at 6.

First, OMB should clarify that agencies should narrowly construe their ability to waive compliance with the minimum practices. For example, the drafted text suggests that if fulfilling the requirement would “increase risks to safety or rights overall,” then CAIOs may waive compliance. OMB should clarify that when it uses the term “risks to safety or rights,” it is specifically referring to the aforementioned purposes that are presumed to be safety-impacting or rights-impacting, and not more generally referring to safety or rights. As currently drafted, agencies may misunderstand the relevant analysis.

Second, OMB should clarify what it expects to be contained within the system-specific risk assessments. In particular, for rights-impacting systems, OMB should clarify that CAIOs must specifically describe how complying with the minimum requirements would “increase risks to rights.” OMB should consider requirements that the Department of Justice’s Civil Rights Division and relevant civil rights officials in each agency be consulted when a CAIO seeks a waiver because compliance with the minimum practices would increase risks to rights. Rarely, if ever, should it be possible for an agency to claim that the very act of documentation, testing, evaluation, ongoing monitoring, and risk mitigation — steps that by their nature are designed to protect rights — would somehow increase risks to rights.

Third, OMB should provide clear examples of what it means for fulfillment of the minimum practices to create “an unacceptable impediment to critical agency operations.”²⁵ As drafted, CAIOs appear to retain sole authority and discretion to determine that abiding by the minimum practices would impede “critical agency operations” and to determine what those specific operations are. If CAIOs take an expansive view of what constitutes an “unacceptable impediment to critical agency operations,” this exception would swallow the rule. OMB could elaborate that impeding critical agency operations means such significant and extraordinary diversion of staff time and resources that the agency risks being unable to fulfill its core mission for the American people. OMB should expect that some agencies may have to divert some staff capacity and resources to ensure compliance with the minimum practices. That fact alone cannot constitute “an unacceptable impediment to critical agency operations.”

²⁵ Id., Section 5(c)(iii) at 14.

4. The final memorandum should ensure that agencies clearly document their anti-discrimination testing process and efforts. It should also require public reporting of these efforts in the AI use case inventories.

As currently drafted, the memorandum suggests that “[a]gencies must document their implementation of these practices and be prepared to report them to OMB, either as a component of the annual AI use case inventory ... or on request as determined by OMB.”²⁶ Separately, the draft memorandum says that starting in 2024 “agencies will be required ... to identify and report additional detail on how they are using safety-impacting and rights-impacting AI” and “how they are managing those risks.”²⁷

It is critical that the final memorandum requires agencies to document their implementation of the minimum practices, so agencies can actually receive effective, constructive feedback, which agencies are required to solicit from “affected groups, including underserved communities, in the design, development, and use of the AI.”²⁸ Such a provision is important: agencies should receive ongoing feedback — through public listening sessions, public hearings, formal comments, and more — from affected communities regarding their use of algorithmic systems. But without transparent documentation as to the choices made when developing and using those systems, as well as in assessing and mitigating disparate impact of those systems, it will be difficult for feedback from affected groups to be effective.

Specifically, it is important that the final memorandum require agencies to clearly document how they approached relevant anti-discrimination testing of algorithmic systems and document how they searched for less discriminatory algorithms. Inherent to this process is a determination that sufficient mitigation of algorithmic discrimination is possible. When an agency identifies that an algorithmic system has disparities and discovers a method to mitigate that discrimination, it should clearly document why they believe that mitigation is sufficient to continue use of the system, and receive feedback from affected groups if they believe that mitigation is sufficient. Similarly, when an agency identifies that an algorithmic system “materially relies on information about a class protected by Federal nondiscrimination laws in a way that could result in algorithmic discrimination or bias against that protected class,” it must “cease the use of the information before using the AI for decision-making.”²⁹ This inherently requires a

²⁶ Id., Section 5(c) at 13.

²⁷ Id., Section 3(a)(iv) at 4.

²⁸ Id., Section 5(c)(v)(B) at 19.

²⁹ Id., Section 5(c)(iv)(A)(1) at 15.

determination as to when a system materially relies on proxies for a protected class. That determination should be documented and justified.

Ultimately, the final memorandum should ensure that future AI use case inventories describe these efforts or ensure that agencies otherwise make this documentation publicly available in an accessible format.

We welcome further conversations on these important issues. If you have any questions, please contact Logan Koepke (Project Director, logan@upturn.org) and Harlan Yu (Executive Director, harlan@upturn.org).