

November 30, 2023

Director Shalanda D. Young
Office of Management and Budget
725 17th St., NW
Washington, D.C. 20503

RE: OMB–2023–0020, Proposed Memorandum for the Heads of Executive Departments and Agencies

Dear Director Young and OMB Colleagues:

We are writing in response to the request for comment on the proposed draft memorandum titled “[Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence](#)” by the Office of Management and Budget (the OMB Memo).

I. Summary of Comment

We applaud the OMB Memo as an important document that generally reflects a thoughtful approach to balancing the benefits of AI innovation with responsible safeguards. This is a critical time to get government technology policy right, particularly when it comes to complex benefits systems that have been challenging to modernize and touch on the lives of millions of Americans. We believe the OMB Memo takes an important step forward to ensure that government agencies can deploy, utilize, and acquire AI in a responsible fashion.

Some of the components of the OMB Memo that we view as particularly important include:

- **Chief AI Officer:** We support the requirement to designate a Chief AI Officer (CAIO), and caution that these leaders will need resources, authority, and people in addition to the title.
- **Hiring:** We applaud the hiring surge, and urge that equal attention be given to upskilling civil servants in their understanding of AI, and the use of academic-agency partnerships.
- **Real-World Testing:** We support the requirement for “[t]esting AI for performance in real-world context,” and caution that current models for technology investment, which support acquisition as a one-time event at the expense of continual development and improvement, are poorly suited for this need.

- **Resources:** We support the call for “[p]roviding or requesting resources via the budget process,” and hope that the majority of those resources go towards internal talent, as government’s over-reliance on vendors will be particularly problematic in the age of AI.

We write, however, to explain why some of the one-size-fits-all “minimum” procedures and practices – applied to all “government benefits or services” programs – may have negative unintended consequences, as the rules intended for higher risk applications will apply to many uses likely to offer substantial benefit and very little risk to the public. Without further clarification from OMB and a clear mandate to tailor procedures to risks, agencies could find themselves tied up in red tape when trying to take advantage of non-controversial and increasingly commodity uses of AI, further widening the gap between public and private sector capabilities. In this letter, we offer some suggestions for distinguishing between these cases and adjusting the guidance to reduce the burden while maintaining AI safety.

Some of the perspectives guiding our recommendations include:

- **Agility and Flexibility:** Because responsible AI requires learning and testing in the real world, the most sensible approach to responsible AI in government calls for agility, not red tape. Adding to what are already years-long delays in the adoption of technology will get us neither the benefits nor the protections the American public deserves.
- **Distinguishing Types of AI and Benefits:** Not all AI is the same. Nor are all benefits systems the same. The risks and rewards of various uses must be assessed distinctly for the type of AI and type of benefit program.
- **The Need for Modernization:** While great caution must be taken with the systems deployed to help Americans most in need, these are also often the systems in greatest need of modernization and innovation, and these considerations must be balanced with the needs of the public in mind.

Our suggestions for further clarification and revision of guidance include:

- **Clarifying the Definition of “Rights Impacting”:** The definition of “rights-impacting” – which triggers “minimum” practices for AI in any “government benefits or services” program – should be clarified and ultimately narrowed to avoid impeding important modernization efforts, especially of the systems most in need of innovation and the systems that support those Americans in greatest need.
- **Distinguishing Among Types of AI:** AI embedded in business processes, such as optical character recognition (OCR) to digitize documents or improved search to retrieve evidence, should be treated differently than AI used to make determinations. OMB should not require agencies to go through the entire software system just to have a human check on low-risk tools that actually improve government services.
 - The extensive processes required to deploy “rights-impacting AI,” especially in benefit programs, should not be triggered for these lower-level, lower-risk uses.
 - OMB must account for cases in which providing for special “AI appeals” could bring a system to its knees. As an extreme example, but that falls potentially under OMB’s interpretation, the U.S. Postal Service cannot enable individuals to opt out of digit recognition – pioneered in 1965 – to send a piece of mail.

- **Avoiding Duplication in Human Review:** The requirement for human review to appeal AI-enabled decisions should not duplicate existing human review mechanisms.
- **Using the Right Baselines:** Agencies should have flexibility to deploy AI tools where algorithmic bias or disparate impacts are lower in magnitude than existing baselines.
- **Securing Data for Equity:** The federal government needs to continue efforts to enable disparity (or equity) assessments. In practice, these have proven extremely challenging for government agencies, as they often lack access to demographic data. Secure, privacy-preserving access to this data and building the competencies for conducting these assessments will be important. OIRA should be engaged to encourage streamlined application of the Paperwork Reduction Act, so that it does not contribute to this burden.
- **Enhancing Public Consultation Without Unduly Delaying AI Policy:** Public consultation on a critical topic such as government use and regulation of AI is a valuable resource to help key government decisionmakers benefit from specialized knowledge and enhance the legitimacy of their decisions while the process also serves to educate the public. That said, it is worth bearing in mind that notice and comment rulemaking can take *years*, and guidance around public consultation should be clarified to reconcile the value of consultation with support for the agile adoption of AI. To thread the needle, agencies can build in consultation mechanisms that are distinct from rulemaking or supplement but do not lengthen particular rulemaking proceedings.
- **Narrowing Consultation by “Affected Groups”:** Consultation by “affected groups” should not be required for development of all AI products regardless of the function. The definition of “affected groups” should be clarified to deter capture by special interests, to avoid unnecessarily triggering notice-and-comment procedures, and to avoid otherwise unnecessarily burdening agencies for no benefit.
- **Adding Nuance to Language on Negative Feedback About AI Deployment:** Language that implies that “agencies must consider not deploying the AI” upon receipt of “negative feedback” is likely to be interpreted strictly, resulting in agencies taking any negative feedback as a trigger to halt, regardless of context. Instead, agencies should be encouraged to take feedback seriously without suggesting that any feedback critical of AI deployment should serve to veto otherwise sensible plans to deploy suitable AI systems.
- **Leveraging Multi-Stakeholder Processes:** In both international and domestic settings, civil servants have successfully used appropriate multi-stakeholder processes to engage key stakeholders without paralyzing policy implementation. We urge OMB to provide guidance that enables, encourages, and empowers agencies to adopt agile, multi-stakeholder design processes.
- **Feasible Notice and Explanation:** Requirements for notice and explanation should not require extensive, individualized explanations of the role of every AI system in each decision that may be technically infeasible or uninformative.

Especially within complex, high-volume benefits systems, we urge OMB to adjust requirements that risk depriving Americans from precisely the kinds of benefits systems they deserve.

We submit the following letter with gratitude and respect for the work of OMB and all the agencies and components who contributed to it.

II. Background

We are individuals who have spent decades working on government technology, both from within government and outside of government. Two of us (Cuéllar and Ho) were Principal Investigators on the report to the Administrative Conference of the United States on the use of artificial intelligence (AI) by federal regulatory agencies (Engstrom et al. 2020). One of us (Cuéllar) served as Justice on the Supreme Court of California, as author of one of the most widely adopted casebooks on administrative law (Mashaw et al. 2020), as a member of the National Academy of Sciences Committee on Responsible Computing Research Ethics and Governance, and worked on regulatory policy in two Presidential administrations, including in the Domestic Policy Council in the White House. One of us (Ho) wrote the White Paper for the National AI Research Resource (Ho et al. 2021) and serves on the National AI Advisory Committee (NAIAC), as Senior Advisor on Responsible AI to the U.S. Department of Labor, and as Director of the Stanford RegLab, which partners with a wide range of government agencies around data science and machine learning. One of us (Pahlka) served as the U.S. Deputy Chief Technology Officer, helping to found the United States Digital Service, was the founder of Code for America to bring 21st century technology to government, and wrote about the urgent need for government to get technology right to serve Americans (Pahlka 2023). One of us (Perez) served in the U.S. Department of Labor in the Office of Unemployment Insurance and conducted extensive work on process automation and improvement in the state of Colorado. One of us (Rodolfa) served as Director of Digital Analytics for the White House Office of Digital Strategy, co-created the Machine Learning for Public Policy Lab at Carnegie Mellon, and has engaged with a wide range of agencies, most recently as Research Director of the Stanford RegLab. One of us (Ray) served for over three decades in the Social Security Administration, including as the deputy executive director of the Office of Appellate Operations, winning the W. Edwards Deming Award for enhancing quality in operations twice, and has written broadly about the need for data-driven modernization of benefits systems (Glaze et al. 2022; Ray and Lubbers 2014; Ray and Sklar 2019; Bajandas and Ray 2018) We write in our individual capacity.

We believe that this is a critical, precedent-setting moment for the U.S. government to lay the foundation for the effective and appropriate use of new AI technologies for the benefit of the country. Government benefits systems, for example, touch the lives of the vast majority of Americans, but rely on procedures and technologies that are sorely in need for innovation and modernization. Roughly [66M Americans](#) – one in every five residents – rely on Social Security benefits. Some [5.4M veterans](#) depend on disability benefits for service-related conditions. And some [46M Americans](#) – representing one of every four workers – are estimated to have relied on unemployment insurance (UI) benefits during the pandemic. But outdated systems, crushing caseloads, and intensive manual review can cause painful delays, alarmingly high rates of error, and lead these systems to collapse right when needed the most. For instance, while UI benefits were paid in a timely fashion for 97% of applications prior to the COVID-19 pandemic, the timeliness rate dropped [below 60%](#) as the crisis caused applications to spike. Human-based decisions of eligibility for the Supplemental Nutrition Assistance Program exhibit errors [44% of the time](#). Such systemic challenges are pervasive across government benefits systems (Ames et al. 2020; Pahlka 2023).

Technology is no silver bullet. But modernization of and innovation within these systems is critical, including the potential use of machine learning and data-driven systems (Ray and Lubbers 2014; Glaze et al. 2022; Administrative Conference of the United States 2022). Existing innovation often happens in

spite of high levels of risk-aversion and inertia at agencies and civil servants do heroic work engaging in path-breaking innovation under these circumstances. Demonstration projects have shown the strong potential for such systems to improve the efficiency and accuracy of benefits determinations, such as through improved search at the Patent and Trademark Office, natural language processing to spot errors in draft decisions at the Social Security Administration, and assistance with navigating complex medical records at the Department of Labor (Engstrom et al. 2020). As the [Administrative Conference of the United States](#) formally recommended, “Agencies, particularly those with large caseloads, should consider whether to use artificial intelligence (AI) tools to help quality assurance personnel identify potential errors or other quality issues.”

We applaud OMB’s Memo as an important document to balance the benefits of AI innovation with responsible safeguards, and are grateful for the opportunity to comment on the draft guidance. In this letter, we first highlight some components of the OMB Memo that we view as particularly important. We then spell out some of what should be clarified or revised to ensure that agencies have the flexibility, resources, and leadership to advance technology in an impactful and responsible fashion.

Our most important caution is that the OMB Memo should clarify that “rights-impacting AI” cannot be interpreted to trigger extensive “minimum” processes and requirements (i.e., *human review* separate from existing review and appeals processes, *opting out* of AI systems when such systems may be embedded in complex layers of benefits systems, *public consultation* for all AI systems pertaining to government benefits systems) for any AI-based output affecting government benefits. We show below that there are a myriad of ways in which AI can be embedded – from optical character recognition (OCR) to digitize documents to improved search to retrieve evidence to smart queuing for improved case processing and error detection of draft decisions – that should not uniformly trigger such extensive processes. Otherwise, this interpretation would deter and ossify innovation within complex, high-volume benefits systems, depriving Americans from precisely the kinds of benefits systems they deserve.

III. Detailed Comments

A. What the OMB Memo Gets Right

We applaud OMB for the extensive work, care, and thoughtfulness that went into this draft memorandum. Overall, the memorandum provides a critical step forward to enabling agencies to utilize technology while also ensuring that its risks are managed.

We highlight six elements that strike us as quite important about the OMB Memo.

1. Chief AI Officer

First, we support the calls for leadership attention to AI through the designation of a Chief AI Officer (CAIO). This follows the [recommendation of the NAIAC](#), as well as recommendations made by [Dr. Lynne Parker](#), the former Director of the National AI Initiative Office, and one of us in [testimony](#) to the Senate Committee on Homeland Security and Governmental Affairs. That said, as noted in those recommendations, the mere designation, whether of an existing official or a new one, is insufficient.

CAIOs will require resources, authority, and people to implement the ambitious agenda set forth both in the OMB Memo, the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Executive Order 13,960, Executive Order 13,859, and related legislative provisions. Some agencies may struggle to identify an appropriate CAIO within 60 days, given the need to onboard individuals, and we believe that flexibility is warranted here to ensure that the most qualified individuals can be hired into the position.

AI Governance Boards: Second, the mandate for the creation of AI Governance Boards strikes us as important to ensure that there is “[a]dequate attention from the agency’s senior leadership.” Responsible AI development, procurement, and governance requires multi-stakeholder teams that can blend technical and domain knowledge. To that effect, we believe the AI Governance Board provisions would be strengthened by (a) expressly including operations staff, who may be best situated to identify where AI adoption can benefit core operations, and (b) enabling outsiders to sit on the AI Governance Board. The first refinement follows the insight from existing innovation teams that leveraged the domain expertise of operations staff to build out “easy wins” for machine learning, such as at the Social Security Administration (Glaze et al. 2022). The second refinement builds on the fact that a number of governance boards have already recruited outside experts, given the lack of resident expertise across a range of subjects (e.g., technical expertise on generative AI, sociotechnical expertise on evaluation, data ethics). Enabling such experts to join as Special Government Employees, under the Intergovernmental Personnel Act, or as detailees could dramatically strengthen governance efforts, given uncertainty around the timeframe and scale of the AI hiring surge.

Agency-Wide Strategic AI Planning: Third, we concur with the OMB Memo’s emphasis on agency-wide strategic planning around AI. Earlier work by Stanford’s RegLab demonstrated inconsistent implementation of the two prior AI Executive Orders (13,960 and 13,859) and the AI in Government Act (Lawrence, Cui, and Ho 2023; 2022), pointing to the need for such strategic planning. The OMB Memo’s emphasis on IT infrastructure, a comprehensive data strategy, managing information security (“authority to operate”) to enable integration of modern software stacks, and the emphasis on hiring technical talent are particularly notable.

Two elements are worth emphasizing further: the availability of the National AI Research Resource (NAIRR), which will also be piloted under the recent AI EO, for participating agencies. Access to such a computing and data resource could be invaluable for agencies that may find the compute and data infrastructure for AI challenging to manage. In addition, we could not be more enthusiastic about the call for digital talent within government. The “AI hiring surge,” however, will not be successful if it does not upskill, develop, and augment civil servants within agencies. The AI Training Act and subsequent implementation of [federal training](#) for acquisition officials by the General Services Administration and OMB are important steps, but such programs should be extended and embedded programmatically with much larger parts of the civil service to explore, understand, and develop competencies around AI. At the same time, we also believe that in addition to Presidential Innovation Fellows and the U.S. Digital Corps, academic-agency partnerships are particularly fruitful ways to build direct pathways from science to impact (Cody and Ho 2022; Maldonado et al. 2022; Engstrom et al. 2020). The Department of Labor, for instance, has [partnered with the RegLab](#) to explore the use of AI in benefits examination; the Environmental Protection Agency partnered with the University of Chicago for environmental

inspections; and Carnegie Mellon partnered with the Department of Homeland Security around cybersecurity. Such partnerships can be important ways to build out capacity.

Real-World Testing: Fourth, we believe one of the most important elements of the OMB Memo lies in its requirement for “[t]est[ing] the AI for performance in a real-world context.” As AI moves into real-world deployment settings, such as in government programs, conventional laboratory benchmarks may be ill-suited to assess how AI will operate in practice. An AI system developed on one university’s hospital patient data may exhibit radical performance drops if deployed to a different hospital setting or as patient profiles change over time (Wu et al. 2021). And because most government AI systems will integrate with human systems, such human interactions are critical to assess. Humans may over-rely, ignore, or veto forms of AI assistance (De-Arteaga, Fogliato, and Chouldechova 2020), each of which may have complicated effects on actual outcomes of interest, and which can only be assessed by testing in real-world contexts. Historically, many government technology efforts, particularly in the procurement setting, have focused efforts upfront: projects are funded for development through extensive ex ante specifications of requirements, and then launched and maintained. But the adaptability of AI is precisely what requires organizations to shift away from this paradigm towards an ethos of continually testing, adjusting, and improving. In partnerships with the U.S. Department of Labor, this has meant comparing how the adoption of an AI system integrates claims examiners via a (randomized) AB test and developing frameworks for how to do so on an ongoing basis. Rigorous evaluations of AI systems require evidence-building and hence have a close synergy with efforts under the Evidence Act, and testing for real world performance will be a central element for trustworthy adoption.

2. Procurement Recommendations

Fifth, we also applaud the OMB Memo’s inclusion of procurement recommendations. The NAIAC has recently issued high level [recommendations](#) around procurement practices under the Federal Acquisition Regulation (FAR), which could be readily integrated as practices that enable improved market research and evaluation of AI solutions. Particularly worth highlighting in the OMB memo is the guidance for agencies to treat data as a critical asset and ensure that contracts retain for the Government sufficient rights to data. Timely, comprehensive, and programmatic access to data relating to all aspects of an agency’s operations is a critical prerequisite for realizing the benefits of AI. Likewise, as the tools of data science and AI become more embedded in an agency’s core operations, limited or restricted access to data poses a significant risk of stifling competition and innovations by drastically increasing the costs of migrating between vendors or building new tools outside of a current vendor’s ecosystem. Similarly, OMB should consider guidance for ensuring that vendors provide to agencies sufficient transparency about their systems to ensure the ability to monitor, assess, and comply with responsible AI requirements. For instance, this may include making available source code, model weights, training data, or other model artifacts to inform agencies' responsible AI assessments. While intellectual property protection can be important, the inability to examine systems can often pose significant barriers to responsible oversight and clear guidance from OMB can help agencies ensure the necessary access to technical materials is reflected in their contracts.

3. Leveraging the Budget Process to Request Needed Resources

Last, we very much support the OMB Memo’s call for “[p]roviding or requesting resources via the the budget process.” This need for resources echoes recommendations by the [NAIAC](#), [ACUS](#), and in [Senate testimony](#). Agencies have an urgent need to develop internal talent for oversight, governance, and modernization. Without federal staff, we are particularly concerned that the federal government may continue to over-rely on vendors to perform core forms of technology modernization, without sufficient internal staff to design, pilot, and evaluate solutions with an intimate understanding of the domain. At the very moment that AI and digital services move closer to core agency operations, it is the time to rethink, as one of us has argued, the conventional distinction between “commercial” and “inherently governmental functions” (Pahlka 2023). The AI revolution should not be outsourced alone.

B. Issues That Require Clarification or Modification

That said, we feel obliged to comment on several items that are in need of more substantial clarification or modification. We focus in particular on the memo’s set of “minimum practices for rights-impacting AI.” These “minimum practices” include a wide-ranging set of procedures, such as an AI impact assessment, ongoing monitoring, public documentation, bias assessment and mitigation, public consultation with affected groups, notice to adversely affected individuals of the use of AI potentially coupled with “explanations” of the AI decision, the provision of a mechanism to “appeal or contest the AI’s negative impacts,” and a mechanism for individuals to “to opt in favor of a human alternative.”

1. Tailor Rights-Impacting AI to Risk of AI and Benefits Program

Because the definition of “rights-impacting AI” can conceivably swallow up most applications of AI, we urge OMB to clarify and right-size this definition. The Memo defines an AI system as “rights-impacting” if it the “output serves as a basis for decision or action that has a legal, material, or similarly significant effect” on rights, such as “[a]ccess to . . . government benefits or privileges.” AI systems are presumed to be rights-impacting if AI is “used to control or meaningfully influence the outcomes” in enumerated domains, but one of the enumerated domains is exceptionally broad, covering “[d]ecisions regarding access to, eligibility for, or revocation of government benefits or services.” Government benefits and services, in turn, are nowhere defined.

This definition of “rights impacting AI” scopes extremely broadly and triggers a wide range of procedures – above and beyond the Administrative Procedure Act – that could threaten core operations across a wide range of government programs. We are concerned that these provisions are not tailored to the reality of modern benefits systems, which process extremely high volumes of cases and where the need for human review and appeals processing has itself created a crisis of due process (Ames et al. 2020). These are systems that are in dire need of innovation and modernization. Requiring special processes for AI innovation – notably public consultation, notice and explanation of AI systems, special appeals systems for AI, and human opt-outs – can seriously impede important modernization efforts. We hence suggest the following revisions:

- The definition should be changed from “serves as a basis for” to “serves as *the principal* basis for”
- Section 5(b)(ii) should be changed to “control or *substantially* influence”

- The enumerated domain in Section 5(b)(J) should be limited to “*significant* government benefits or services” or benefits and services should be explicitly defined to encompass a narrower set of high-risk benefits and service programs.

Our reasons are fourfold. First, absent these revisions, the definition could implicate a vast range of government benefits. These include:

- The issuance of campsite permits by the National Park Service,
- The decision by a Postal Service clerk on what to charge to mail a package,
- Grant decisions by the National Science Foundation for award amounts as low as a few thousand to tens of millions of dollars for roughly 40,000 applications annually,
- Inspections of drug manufacturing plants, farms, and food imports by the Food and Drug Administration,
- Patent examination decisions for over 600,000 annual applications by the Patent and Trademark Office,
- Determinations of eligibility for unemployment insurance for over 200,000 applications each week,
- Audit decisions for over 700,000 annual audits by the Internal Revenue Service.

The sheer volume, heterogeneity, and diversity of government benefit programs defies simplistic application of uniform procedures. Or, as put by administrative law scholar [Prof. Asimow](#): there is “no uniformity of a sort that can be easily governed by [] trans-substantive” provisions for these kinds of determinations.

Second, within a given domain, AI can be deployed in a wide range of ways (Engstrom et al. 2020; Cuéllar and Huq 2022). Many of the most promising uses of AI are lower-level forms of process automation or assistive technology, where humans continue to issue final decisions. The Social Security Administration, for instance, might employ AI to improve disability appeals processing by: (a) using AI to improve OCR to digitize documents; (b) aiding in the summarization, search, or translation of documents; (c) clustering solely to determine the order in which cases are decided by an adjudicator; (d) providing a check for potential errors in draft decisions with natural language processing (Glaze et al. 2022). As of 2023, the Internal Revenue Service required individuals to [manually enter digits](#) on paper tax returns, leading to a backlog of 10.2M unprocessed returns, when digit recognition has been deemed a solved problem by AI for decades (LeCun et al. 1998). (For reference, the U.S. Postal Service began to adopt OCR in 1965.) Treating all of these potential applications as equally “rights-impacting” makes little sense. Although the OMB memo describes a mechanism for the CAIO to exempt specific applications from the minimum practices triggered by “rights-impacting” applications, we have significant concerns with an orientation that starts from an overly-broad presumption of high risk and requires proactive action to identify exceptions. Beyond the unnecessary process involved, the incentives in a policy implementation context are generally poorly aligned with granting such exceptions as acting to waive requirements may carry a risk of scrutiny or sanctions. Such dynamics driven by risk-aversion are unfortunately all too common in the administrative state, particularly where technology is involved, leading to adverse outcomes such as outdated and ineffective software being deployed on modern GPS satellites based on little more than an off-hand example in policy guidance that was misinterpreted as a requirement (Pahlka 2023).

Third, many of these systems already have human review and appeals mechanisms, and the Memo should clarify that any rights to contest AI would not duplicate, and instead simply be folded into, such existing human review mechanisms. While it may be true that such processes might need to adapt to the nuances of AI systems — for instance, to guard against automation biases and over-reliance on the initial AI determinations in the appeal outcome itself — appeals of AI-informed and entirely human-derived decisions are not inherently different and should not require independent lanes. Further, it makes no sense for claimants to be able to opt out of AI assistance that sits layers within the internal process of the agency — such as AI assisted OCR, document summarization, and assistance — as this could prove technically infeasible and procedurally laborious. Nor would it make sense to provide for special “AI appeals” in instances where appeals could bring a system to its knees. It would be absurd, for instance, to require the U.S. Postal Service to enable individuals to opt out of digit recognition to send a piece of mail. Such manual review would cause undue budget pressure and could threaten the health of the postal system as a whole.

Fourth, our requested revisions clarify that process should be triggered only in instances where AI impacts are substantial. The OMB Memo should not require looking down the entire software stack to require human review of low-risk tools that improve government services. Instead, OMB should enable agencies to make determinations and focus their assessment efforts on AI tools that are *substantial* in effect. Indeed, this is the essence of how the Administrative Procedure Act protects agency discretion for general statements of policy or rules of agency organizations, procedure, or practice (Engstrom and Ho 2020). In many ways, OMB’s binary rule — if (1) AI is rights-impacting, then (2) human review, opt-out, appeals, and public hearing are triggered — neglects the overarching lesson from procedural due process, which is needed to tailor process to risk. See *Mathews v. Eldridge*, 424 U.S. 319 (1976).

2. Clarify Obligations to Assess Algorithmic Discrimination

We applaud the Memo’s requirement to proactively identify and mitigate algorithmic bias and disparate impacts. We urge clarification on two issues.

First, it is worth noting how deeply challenging disparity assessments have proven for government agencies. Many agencies lack access to demographic data that would enable such assessments. Twenty-one of 25 agencies filing equity action plans have noted serious data challenges to conduct the required equity assessment under Executive Order 13,985 (Gupta et al. 2023). We urge the administration to continue work to enable agencies to obtain secure, privacy-preserving access to be able to conduct such assessments. We speak from personal experience, with one of us having worked on the team to conduct the assessment of racial disparities in tax audits, which was a substantial effort (Elzayn et al. 2023). Such assessments can be substantial undertakings that require OMB to cooperate in other regards, such as clearance for data collection under the Paperwork Reduction Act (PRA) and improved guidance under the Privacy Act of 1974 to enable secure, privacy-protecting linkage to administrative records. Just as the Memo acknowledges in footnote 33 that the PRA should not prohibit input from stakeholder consultation, the PRA should not stand in the way of collecting demographic data necessary to conduct disparity assessments.

Second, we urge OMB to revise the draft language to account for the fact that AI tools may exhibit disparities – but ones that can be lower in magnitude than existing baselines (Kleinberg et al. 2019). Currently, the Memo provides that if “there are significant disparities [and mitigation is not possible], then agencies should not use or integrate the AI tool.” This neglects the fact that existing baselines may exhibit greater disparities than the AI tool. As the IRS example illustrates, it may not be possible to reduce disparities to zero (and naively doing so itself poses challenges under antidiscrimination law (Xiang 2021; Xiang and Raji 2019; Ho and Xiang 2020)), but if the AI system reduces disparities significantly relative to baselines, there may still be a strong case for adoption (Coglianese and Lai 2021). We suggest changing this language to: “If adequate mitigation of the disparity is not possible *and the AI system does not reduce disparities relative to existing baselines*, then agencies should not use or integrate the AI tool.”

3. Public Consultation

The OMB Memo provides that, “[t]o the extent practicable . . . agencies must consult affected groups” and that, “[i]n the event of negative feedback, agencies must consider not deploying the AI.” We ourselves have seen extensive value to consultation, iteration, and design processes that enable a wide range of stakeholders to contribute to technology. Consultation, feedback, and user-oriented design can be of immense value to responsible development (Lu et al. 2022). Specifically, such consultation can help key government decisionmakers benefit from specialized knowledge and enhance the legitimacy of their decisions while the process also serves to educate the public.

That said, participatory design principles in AI, particularly within the public sector are far from settled. The Memo alludes to notice-and-comment rulemaking and public hearings, but such proceedings can be tremendously burdensome for agencies, with disparities in actual participation. As the [Government Accountability Office](#) reported, “[b]ased on the limited information available, the average time needed to complete a [major] rulemaking across our 16 case-study rules was about *4 years*” (emphasis added).¹ What is needed for responsible agency adoption of AI is *agility*, not burdensome processes, especially in light of the huge variety of government programs and AI systems.

In practice, these valuable user-centered development techniques that can support rapid iteration and continual improvement are too often stifled by cumbersome and lengthy processes of ensuring compliance with the PRA or undergoing notice and comment rulemaking (Pahlka 2023). Footnote 33 in the OMB Memo briefly acknowledges this potential barrier, but we would strongly encourage OMB to better clarify this point and provide direct, affirmative, and concrete guidance to agencies on how they can engage with stakeholders throughout the AI project lifecycle in an effective and timely manner.

We urge several other clarifications. First, OMB should carefully think about triggering process when none is required under the Administrative Procedure Act (APA). The APA exempts rules of internal agency organization and guidance documents from requirements of public consultation. And there is good reason for these exemptions: for the same reason that “ordinary business operations” are exempt from shareholder votes for public corporations, agencies must retain discretion to internally organize, prioritize,

¹ Even those who dispute that rulemaking is “ossified” estimate that it takes agencies *1-2 years* to complete most rules (Yackee and Yackee 2011). Notice and comment rulemaking, as currently practiced, is not the solution for technology design.

and allocate resources. Many AI systems may fall under these exemptions – e.g., an AI system to assign cases to investigators – and the OMB Memo should not contradict the APA’s protection for agencies to be able to make internal allocation decisions.

Second, OMB should clarify what is meant by “affected groups.” Consultation by affected groups should not be required for development of all AI products regardless of the function. External stakeholders should be consulted for externally facing AI products, such as chatbots or application systems. In other instances, such as enforcement tools, consultation could backfire (e.g., “IRS, please consult with tax cheats”). Other AI products (OCR, document digitization, scheduling, search) that are backend process supports have no direct implications for external groups, nor would such groups understand the end-to-end process to provide meaningful feedback. Adding this requirement in these types of situations would generate work with little value added for already burdened agencies. For many internal facing process improvements, the most important “affected group” are federal employees themselves, as the Memo notes. It will remain critical for the CAIO to discern when consultation is required and to ensure that the extent of consultation is tailored to the risk, scope, and impact of the AI system.

Third, OMB should clarify how agencies are expected to obtain meaningful feedback from individuals. Will agencies be afforded funding to provide stipends or the like to customers who provide input or is feedback solely voluntary? What happens when agencies are unable to secure volunteers or input from union representatives? Will this be added as a job duty for federal staff? Flexibility is required here given that participatory design in the public sector context is still largely uncharted terrain.

Fourth, OMB should change the language that implies that “agencies must consider not deploying the AI” upon receipt of “negative feedback.” The current language can be misinterpreted to imply that any negative feedback must trigger a consideration to not deploy. But with any large group, there may be some forms of negative feedback, even when the overwhelming majority may be positive. We recommend revising this language to require the agency to “consider the implications of feedback on the design, development, and deployment of the AI.”

Overall, we urge OMB to provide guidance that enables, encourages, and empowers agencies to adopt agile, multi-stakeholder design processes. Responsible AI calls for agility, not red tape.

4. Notice and Explanations

The OMB Memo provides that, where practical and lawful, agencies “must notify individuals when AI meaningfully influences the outcome of decisions specifically concerning them, such as the denial of benefits.” In addition, the Memo encourages agencies “to provide *explanations* for such decisions and actions” (emphasis added).

Notice and explanation are, of course, linchpins of due process and administrative law. It is blackletter law that agencies must, by default, provide reasoned explanations for agency actions, including benefits denials. See *Motor Vehicle Manufacturers Association v. State Farm Mutual Automobile Insurance Co.*, 463 U.S. 29 (1983); *Mathews v. Eldridge*, 424 U.S. 319 (1976).

Yet what is concerning about the draft language is that it can be misinterpreted to imply distinct notices and explanations specific to each application of AI systems.

First, while many legislative and regulatory proposals focus on “explainable AI,” it is important to note that the technical, scientific basis for AI explanations is lacking (Amarasinghe et al. 2023; Guha et al. 2024). Put simply, OMB may be requiring something that is not currently technically feasible. Footnote 37 acknowledges that explanation is “often not technically feasible,” but elides over significant tensions in the requirement. “[A]gencies should characterize the *general* nature of such AI decisions,” but a decision-specific “explanation[] should be . . . meaningful, useful, and as simply stated as possible.” Documenting an information retrieval system may meet the former, but not the latter. Systems-level disclosures are, of course, technically feasible and may also be more likely to uncover design and development problems.

Second, it is unclear what AI-specific harms or risks OMB aims to address with these requirements, relative to existing requirements for transparency. What benefit denials are currently not already subject to forms of notice and explanation? Is the explanation intended to require an explanation of the AI system, the AI output, or something else? When notice and explanation are already required, OMB needs to provide more guidance on what, if anything, should be provided for each decision that is specific to the AI system. An administrative law judge, for instance, must provide an explanation for reaching a decision. But that should not necessarily require the administrative law judge to explain to each litigant, for instance, the search algorithm for precedents that could be embedded legal research software, the case queuing mechanism to assign cases to adjudicators, or automated guidance that may suggest materials to consult.

Third, OMB should clarify what it means to “meaningfully influence[] the outcome.” In many instances, a decision may be the same, regardless of whether processed by a human or an AI system – e.g., benefit amount calculations or employment requalification requirements. When a human utilizes AI output as part of the decision process, when does that trigger this additional explanation requirement? Is the influence of the AI system what matters? When an AI system improves the accuracy of determinations (e.g., Ray and Lubbers 2014), is then a *lesser* level of explanation required for erroneous human decisions (because AI “meaningfully influence[d] the outcome” relative to the human baseline)? This would be an odd one-sided explanation requirement, and AI-specific requirements should not deter from investing in the continuous improvement of government systems. Per OMB’s laudable requirement to test AI for performance in the real-world context, we should measure, evaluate, and adopt systems that improve complex government systems relative to the baseline.

Last, while we very much favor transparency of technical systems, as we have noted above, there is a long line of research that establishes that individual notices and appeals can be exceptionally poor at correcting systemic errors in benefits programs (Mashaw 1973; 1985; Ames et al. 2020; Hausman 2021), with particularly acute problems when it comes to technology (Citron 2007). Due process may be much better served through alternative mechanisms for discovering errors, such as by comparing AI assistance with a representative set of decisions reached without AI assistance (Engstrom and Ho 2020).

* * * *

We admire the accomplishments of the AI Executive Order and strongly support the core tenets of the draft OMB Memo. Because the memo will be central to how hundreds of government agencies pursue technology modernization – and, in turn, the potential use of a wide range of AI and machine learning approaches – we think it is critical that the framework promotes responsible innovation, while paying attention to the breathtaking variety of government programs and AI. A one-size-fits-all procedural solution will not serve the administrative state well and, if misinterpreted, could simply tie agencies up in red tape. By making the above clarifications and ensuring that this is a time for investment in public sector capacity, we believe the OMB Memo can ensure that process is tailored to risk and that government agencies are prepared to responsibly use, develop, acquire, and govern AI.

It is high time for the government to get technology right. With the clarifications and changes suggested above, the OMB Memo will be a pivotal step forward, and we stand ready to support the country in these efforts.

Sincerely,



Mariano-Florentino Cuéllar
President, Carnegie Foundation for International Peace
Former Justice of the Supreme Court of California
Former Stanley Morrison Professor of Law, Stanford Law School



Daniel E. Ho
William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science and Computer Science (by courtesy), Stanford University
Senior Fellow, Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Senior Fellow, Stanford Institute for Economic Policy Research (SIEPR)
Director, Stanford Regulation, Evaluation, and Governance Lab (RegLab)
Member, National AI Advisory Committee (NAIAC)



Jennifer Pahlka
Former US Deputy Chief Technology Officer, White House Office of Science and Technology Policy

Senior Fellow, Federation of American Scientists
Senior Fellow, Niskanen Center
Co-founder and Board Chair, US Digital Response



Amy Perez
Former Timeliness Supervisor
Central Response Division
Office of Unemployment Insurance
U.S. Department of Labor



Kit Rodolfa
Research Director, Stanford Regulation, Evaluation, and Governance Lab (RegLab)
Former Director of Digital Analytics, White House Office of Digital Strategy



Gerald Ray
Former Administrative Appeals Judge
Former Deputy Executive Director, Office of Appellate Operations
Social Security Administration

References

- Administrative Conference of the United States. 2022. “Quality Assurance Systems in Agency Adjudication.” *Federal Register* 87 (January): 1722.
- Amarasinghe, Kasun, Kit T. Rodolfa, Hemank Lamba, and Rayid Ghani. 2023. “Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions.” *Data & Policy* 5: e5. <https://doi.org/10.1017/dap.2023.2>.
- Ames, David, Cassandra Handan-Nader, Daniel E. Ho, and David Marcus. 2020. “Due Process and Mass Adjudication: Crisis and Reform.” *Stan. L. Rev.* 72: 1.
- Bajandas, Felix F., and Gerald K. Ray. 2018. “Implementation and Use of Electronic Case Management Systems in Federal Agency Adjudication,” May.
- Citron, Danielle Keats. 2007. “Technological Due Process.” *Wash. UL Rev.* 85: 1249.
- Cody, Sara H., and Daniel E. Ho. 2022. *How to Build Academic-Public Health Partnerships: The Stanford - Santa Clara County Experience with COVID-19 Response*. Edited by Tamar Bauer. Brookings.
- Coglianesse, Cary, and Alicia Lai. 2021. “Algorithm vs. Algorithm.” *Duke LJ* 71: 1281.
- Cuéllar, Mariano-Florentino, and Aziz Z. Huq. 2022. “Artificially Intelligent Regulation.” *Daedalus* 151 (2): 335–47. https://doi.org/10.1162/daed_a_01920.
- De-Arteaga, Maria, Riccardo Fogliato, and Alexandra Chouldechova. 2020. “A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores.” In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. CHI '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376638>.
- Elzayn, Hadi, Evelyn Smith, Thomas Hertz, Arun Ramesh, Robin Fisher, Daniel E Ho, and Jacob Goldin. 2023. “Measuring and Mitigating Racial Disparities in Tax Audits.” *SIEPR Working Paper*. <https://siepr.stanford.edu/publications/measuring-and-mitigating-racial-disparities-tax-audits>.
- Engstrom, David Freeman, and Daniel E. Ho. 2020. “Algorithmic Accountability in the Regulatory State.” *Yale Journal on Regulation* 37: 800–854.
- Engstrom, David Freeman, Daniel E. Ho, Catherine Sharkey, and Mariano-Florentino Cuéllar. 2020. *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Administrative Conference of the United States.
- Glaze, Kurt, Daniel E. Ho, Gerald K. Ray, and Christine Tsang. 2022. “Artificial Intelligence for Adjudication: The Social Security Administration and AI Governance.” In *The Oxford Handbook of AI Governance*. Oxford University Press, Handbook on AI Governance.
- Guha, Neel, Christie M. Lawrence, Lindsey A. Gailmard, Kit T. Rodolfa, Faiz Surani, Rishi Bommasani, Inioluwa Deborah Raji, Mariano-Florentino Cuellar, Colleen Honigsberg, and Daniel E. Ho. 2024. “AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing.” *George Washington Law Review*.
- Gupta, Arushi, Victor Wu, Helen Webley-Brown, Jennifer King, and Daniel E. Ho. 2023. “The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in US Government.” In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 492–505.
- Hausman, David. 2021. “Reviewing Administrative Review.” *Yale Journal on Regulation* 39: 1059–1123.
- Ho, Daniel E., Jennifer King, Russell Wald, and Christopher Wan. 2021. “Building a National AI Research Resource: A Blueprint for the National Research Cloud.” Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/sites/default/files/2022-01/HAI_NRCR_v17.pdf.
- Ho, Daniel E., and Alice Xiang. 2020. “Affirmative Algorithms: The Legal Grounds for Fairness as Awareness.” *University of Chicago Law Review Online*.

- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass Sunstein. 2019. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis* 10 (April): 113–74.
- Lawrence, Christie, Isaac Cui, and Daniel Ho. 2023. "The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at U.S. Federal Agencies." In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 606–52. AIES '23. Association for Computing Machinery.
- Lawrence, Christie, Isaac Cui, and Daniel E. Ho. 2022. "Implementation Challenges to Three Pillars of America's AI Strategy." Stanford RegLab.
- LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324.
- Lu, Lisa, 'Agostino Alexis D, Sarah L. Rudman, Derek Ouyang, and Daniel E. Ho. 2022. "Designing Accountable Health Care Algorithms: Lessons from Covid-19 Contact Tracing." *NEJM Catalyst* 3 (4): CAT.21.0382. <https://doi.org/10.1056/CAT.21.0382>.
- Maldonado, Peter, Angie Peng, Derek Ouyang, Jenny Suckale, and Daniel E. Ho. 2022. "Science Translation During the COVID-19 Pandemic: An Academic-Public Health Partnership to Assess Capacity Limits in California." *American Journal of Public Health* 112 (2): 308–15. <https://doi.org/10.2105/AJPH.2021.306576>.
- Mashaw, Jerry L. 1973. "Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy Fairness and Timeliness in the Adjudication of Social Welfare Claims." *Cornell L. Rev.* 59: 772.
- . 1985. *Bureaucratic Justice: Managing Social Security Disability Claims*. Yale University Press.
- Mashaw, Jerry L., Richard A. Merrill, Peter M. Shane, M. Elizabeth Magill, Mariano-Florentino Cuellar, and Nicholas R. Parrillo. 2020. *Administrative Law, The American Public Law System, Cases and Materials*. 8th ed. West Academic.
- Pahlka, Jennifer. 2023. *Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better*. Macmillan.
- Ray, Gerald K., and Jeffrey S. Lubbers. 2014. "Government Success Story: How Data Analysis by the Social Security Appeals Council (with a Push from the Administrative Conference of the United States) Is Transforming Social Security Disability Adjudication, A." *Geo. Wash. L. Rev.* 83: 1575.
- Ray, Gerald K., and Glenn Sklar. 2019. "An Operational Approach to Eliminating Backlogs in the Social Security Disability Program." *McCrery-Pomeroy SSDI Solutions Initiative*, June.
- Wu, Eric, Kevin Wu, Roxana Daneshjou, David Ouyang, Daniel E. Ho, and James Zou. 2021. "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals." *Nature Medicine* 27 (4): 582–84.
- Xiang, Alice. 2021. "Reconciling Legal and Technical Approaches to Algorithmic Bias." *Tennessee Law Review* 88 (3).
- Xiang, Alice, and Inioluwa Deborah Raji. 2019. "On the Legal Compatibility of Fairness Definitions." *Workshop on Human-Centric Machine Learning at the 33rd Conference on Neural Information Processing Systems*.
- Yackee, Jason Webb, and Susan Webb Yackee. 2011. "Testing the Ossification Thesis: An Empirical Examination of Federal Regulatory Volume and Speed, 1950-1990." *Geo. Wash. L. Rev.* 80: 1414.