Upturn
Toward Justice
in Technology

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.

Upturn
Toward Justice in Technology

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-administration-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 different 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://ww2.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 Euller, 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/RRA1800/RRA1853-1/RAND_RRA1853-1.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).

TESTIMONY OF JENNIFER PAHLKA, FORMER US DEPUTY CHIEF TECHNOLOGY OFFICER, SENIOR FELLOW, FEDERATION OF AMERICAN SCIENTISTS AND THE NISKANEN CENTER
BEFORE THE COMMITTEE ON HOMELAND SECURITY AND GOVERNMENT AFFAIRS, U.S. SENATE
ON HARNESSING AI TO IMPROVE GOVERNMENT SERVICES AND CUSTOMER EXPERIENCE

JANUARY 10, 2024

Chair Peters, Ranking Member Paul, and members of the Committee, I appreciate you inviting me here today to speak on this critical topic.

How the US government chooses to respond to the changes AI brings is indeed critical, especially in its use to improve government services and customer experience. If the change is going to be for the better (and we can't afford otherwise) it will not be primarily because of how much or how little we constrain AI's use. Constraints are an important conversation, and AI safety experts are better suited to discuss these than me. But we could constrain agencies significantly and still get exactly the bad outcomes that those arguing for risk mitigation want to avoid. We could instead direct agencies to dive headlong into AI solutions, and still fail to get the benefit that the optimists expect. The difference will come down to how much or how little ***capacity and competency*** we have to deploy these technologies thoughtfully.

There are really two ways to build capacity: having more of the right people doing the right things (including but not limited to leveraging technology like AI) and safely reducing the burdens we place on those people. AI, of course, could help reduce those burdens, but not without the workforce we need – one that understands the systems we have today, the policy goals we have set, and the technology we are bringing to bear to achieve those goals. Our biggest priority as a government should be building that capacity, working both sides of that equation (more people, less burden.)

Building that capacity will require bodies like the US Senate to use a wide range of the tools at its disposal to shape our future, and use them in a specific way. Those tools can be used to create mandates and controls on the institutions that deliver for the American people, adding more rules and processes for administrative agencies and others to comply with. Or they can be used to enable these institutions to develop the capacity they so desperately need and to use their judgment in the service of agreed-upon goals, often by asking what mandates and controls might be removed, rather than added. This critical AI moment calls for ***enablement***.

The recent executive order on AI already provides some new controls and safeguards. The order strikes a reasonable balance between encouragement and caution, but I worry that some of its guidance will be applied inappropriately. For example, some government agencies have long been using AI for day to day functions like handwriting recognition on envelopes or improved search to retrieve evidence more easily, and agencies may now subject these benign, low-risk uses to red tape based on the order. Caution is merited in some places, and dangerous in others, where we risk moving backwards, not forward. What we need to navigate these frameworks of safeguard and control are people in agencies who can tell the difference, and who have the authority to act accordingly.

Moreover, in many areas of government service delivery, the status quo is frankly not worth protecting. We understandably want to make sure, for instance, that applicants for government benefits aren't unfairly denied because of bias in algorithms. The reality is that, to take just one benefit, one in six determinations of eligibility for SNAP is substantively incorrect today. If you count procedural errors, the rate is 44%. Worse are the applications and adjudications that haven't been decided at all, the ones sitting in backlogs, causing enormous distress to the public and wasting taxpayer dollars. Poor application of AI in these contexts could indeed make a bad situation worse, but for people who are fed up and just want someone to get back to them about their tax return, their unemployment insurance check, or even their company's permit to build infrastructure, something has to change. We may be able to make progress by applying AI, but not if we double down on the remedies that failed in the Internet Age and hope they somehow work in the age of AI. We must finally commit to the hard work of building digital capacity.

History of Digital Enablement of Services in Government

Customer experience changed dramatically during the Internet era – we no longer wait in line at the bank to deposit a check or at the airport for a taxi. Many of the interactions we used to think of as customer service have disappeared, submerged into a layer of technology and data that answers the questions customer service used to ask. Who are you? Your bank knows. Where are you? Your ride hailing service knows. The public mostly likes these changes, but more importantly, it expects them. It now feels odd, even a little scary, to be asked questions the institution should know the answer to. It's hugely frustrating to wait weeks, even months, for an answer that you know relies on some basic math a computer could do in nanoseconds if it just were just allowed to process the data you have just given it. "This isn't that hard," the veteran says as his application languishes in a backlog. We've made a lot of progress, but we are still struggling to gain the benefits the Internet era offered, (the White House recently wrote that only two percent of government forms are available online!)[1], and the next era is already upon us.

How did we get here? A little history helps explain. Starting as far back as the 1960s, but particularly in the 1990s, when companies like Amazon and Google were emerging, leadership in government (both Democrats and Republicans) mistook what ultimately proved to be a

---

[1] Why the American People Deserve a Digital Government, Clare Martorana, Federal CIO, September 22, 2023
https://www.whitehouse.gov/omb/briefing-room/2023/09/22/why-the-american-people-deserve-a-digital-government/

massive digital revolution for a mere tactical shift in the tools of implementation. Tools are things you buy, so leadership saw digital as a problem of purchasing. Instead of recognizing that no institution, public or private, would be able to operate effectively in the coming decades without basic digital competence, and therefore hiring people who understood this brave new world, our government developed extensive processes and procedures for *buying* digital technology as if it were simply a commodity. Today, as we bemoan the lack of expertise in highly specialized, complex domains like advanced software, it's worth noting that the inner workings of procurement seem as specialized, complex, and mysterious to the layperson as the inner workings of an AI model. Government is clearly capable of developing capacity in specialized domains. We just picked the wrong ones.

We have treated digital much like we treat pens, paper clips, or vehicles that the General Services Administration buys for agencies: we don't need to know how it works, we just need to acquire it. Once we've acquired it, other than perhaps a maintenance contract, we're done. Today, though it takes us a painfully long time to do so, government knows how to acquire static software. What we need to acquire are capabilities.

Flexible Capabilities

Like most of what I will cover today, buying static software like we buy pens or cars was not a good idea in the Internet era, but it is a catastrophically bad one in the AI era. Software systems were always less static than our procurement frameworks allow for, and AI is orders of magnitude more dynamic. AI systems have all the dynamic characteristics of the previous software era, but are literally learning all the time, and therefore constantly changing in ways that we don't entirely understand. Therefore, responsible and effective use of AI *must* involve constant learning and testing in the real world. Academics have shown, for instance, that an AI system developed on one university's hospital patient data can perform radically differently if deployed to a different hospital setting or as patient profiles change over time[2]. Our current "once and done" frameworks don't allow for this ongoing evaluation, and our workforce is not suited to these challenges. We cannot simply engage procurement officers to evaluate and purchase a system like that, and hope it works out. AI demands agility and competence in ways we can no longer afford to ignore.

To illustrate the limitations of our legacy government procurement frameworks, it might be helpful to hear an example of what it's been like for government technologists trying to guide a previous transition: the move to the cloud. One of the early recruits to 18F, Jez Humble, was working with a contracting officer in an attempt to purchase cloud services. Jez had prepared enormous amounts of data in advance of meeting with the contracting officer, but in the meeting, he found that he lacked the one piece of information the officer needed: how much these cloud services would be used. The officer could not put out a bid to procure a service if he didn't know how much he would be asking for.

[2] 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.

One of the key advantages of cloud computing, of course, (though not the only one) is the flexibility it offers. Instead of having to guess how much infrastructure you'll need well ahead of launching, say, a website, and buying what you hope is the right number of servers and the sufficient bandwidth, you can essentially rent a flexible amount of capacity from a cloud computing provider and only pay for what you use. If traffic is less than expected, you save money. If it's more, you pay more, but at least your website stays up as the cloud provider seamlessly handles the extra load. Jez couldn't tell the contracting officer how much "cloud" he needed to buy, because not knowing is exactly the point of this technology. Jez was looking to acquire a cloud capability; contracting could only acquire a fixed, known quantity.

The contracting officer wanted to help Jez, but continued to insist that nothing could move forward without specifying a fixed amount. Jez explained the value of the cloud computing model in every way he knew how. It's a bit like gas for your car, he tried, to no avail. They went back and forth for over two hours. Finally the contracting officer took a deep breath and said, "Let me explain how contracting works in the US government. We put in an order for 100 sandbags, we get 100 sandbags." And the conversation was over.

Jez did ultimately succeed in buying cloud services (at terms far less favorable than the private sector because of government's bespoke needs), but the process took orders of magnitude more effort, time, and money than it would have under a less rigid procurement framework. This rigidity has been a huge hindrance to the ability of government to serve its people; it will be even more obstructive if we hope to use AI. To ease that rigidity, we will need to provide agencies more flexibility, not less. We will need to enable more than we mandate.

Data Ownership

Another example of how procurement will need to change is illustrated by our problems with data ownership. Processes for software acquisition over the past several years, grounded in misguided assumptions about how to evaluate vendors, have failed time and time again to ensure government's access to its own data. Without data, there is no AI. The Office of Management and Budget's (OMB) guidance to agencies about implementing the AI executive order stresses the need to treat data as a critical asset and ensure that contracts retain sufficient rights to data. This is essential moving forward, but there are few government agencies who can confidently say that they have those sufficient rights now, on both a legal and practical basis. (Sometimes, agencies seem to have the appropriate rights on paper, but when it comes to accessing data from their vendor, they find there are barriers, including but not limited to additional, unbudgeted costs.) This is particularly problematic when it comes to the equity audits that are now required for certain uses of AI by the new executive order.[3] The majority of agencies now filing equity action plans lack the data needed to do so, some of which (but not

[3] 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.

all) is due to vendor control. In the meantime, for this reason and many others, adoption of AI to improve services will be stalled.

Once again, we have an issue that's been problematic in the past, but becomes orders of magnitude more problematic in the AI era. Vendors will have even more powerful ways to stifle competition, lock agencies in, and skirt appropriate transparency and oversight unless government finally recognizes the value of its data and moves decisively to retain it. Vendors can be incredibly valuable partners in the mission, but the coming era requires government to step up and create the rules of the road for vendors to follow that truly serve the public.

Reducing Burdens

Jez's experience is also a great example of what I mean when I say that the other half of building capacity is reducing the burden on the people you have. Jez represents exactly the kind of talent we needed (and still need) in the Internet era: expert in the latest technologies, mission-driven, and a creative thinker. His counterparts in AI are the kind of people we seek to recruit today. And we succeeded in getting him to work in government for a time, between his tenure at hot start-ups and companies like Google. But he spent most of this time in government not deploying the latest technologies to improve government services, but fighting bureaucratic and administrative battles. The American people got some fraction of the value we could have had from Jez. We must not only recruit the right people, but do whatever we can to make it possible for them to do the job they came to do.

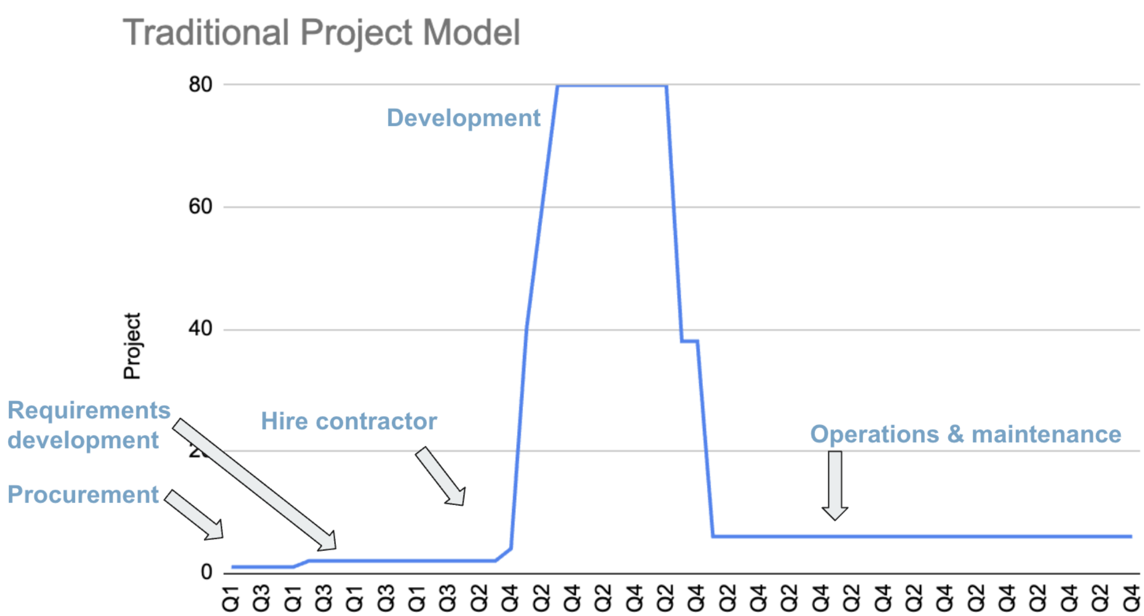
This imperative is not limited to tech workers. To improve customer experience, we will need far more people who understand data and technology. But what the public wants from customer service is answers: Where is my check? Why did I get this IRS notice? If we use AI just to make it easier to talk to government, but not to get those answers, we will fail. The key reason those answers are hard to come by is the enormous complexity of government programs. I worked on California's unemployment insurance crisis during the pandemic, and encountered what is close to 10,000 pages of regulations governing what could be a relatively simple program. A claims processor working with my colleague kept calling himself "the new guy" because he was still learning the ropes. He had been with the agency for *17 years*. But recall that unemployment insurance dates back to the 1935 Social Security Act. We've been adding rules and mandates for close to 90 years now. We almost never remove them. It's no wonder the program is still in peril. It is collapsing under its own weight, weight that federal and state agencies can't shed on their own.

It is tempting to say that AI will help us by understanding those 90 years of accumulated policy cruft for us. This is appealing in the short term and very dangerous in the long term. We can't have a government so complex that one algorithm talks to the other at such a level of complexity that humans are out of the loop. Think about this problem in the context of the Department of Defense, where navigating the complexity and sheer volume of Pentagon policy — equivalent to 100 copies of "War and Peace" — slows everything from acquisitions to hiring to logistics to combat operations. I, for one, am not eager to live in a world where only AIs can tell our uniformed service members when they can and cannot shoot. But we can, and should, use AI to suggest dramatic simplifications to these overwrought frameworks and make those new leaner frameworks the law of the land. The greatest gift this body could give to the agencies and the American public they serve is a massive, thorough decluttering and spring cleaning. AI makes it possible, but only you in this chamber can make it happen.

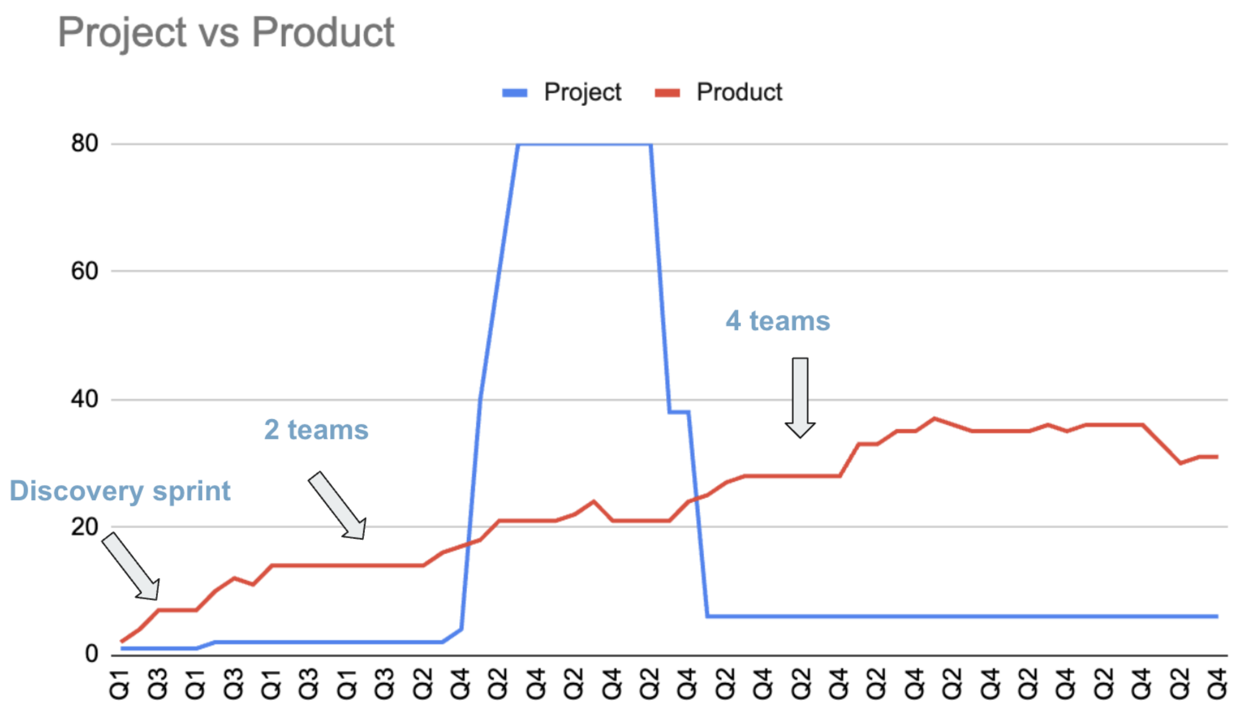
Funding

Government procurement is a poor fit for competence in AI, but funding is upstream of procurement, and equally ill-suited to the task, in similar ways. Not only do we procure software as if it were static, we also fund it that way, and thus make it both worse and more expensive. This is best illustrated through a series of graphs, each one fictional but representative of two fundamentally different approaches to funding software (in both the current and coming paradigms.)

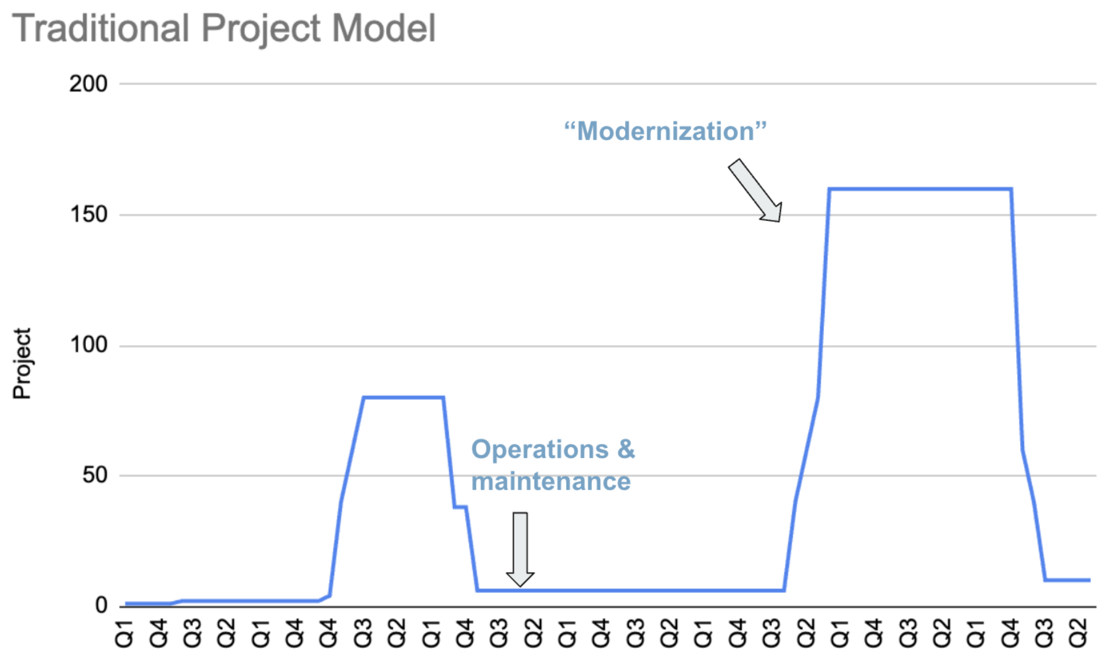
Government follows a “project” model. The following graph shows the number of staff who work on an IT project at its outset, as requirements are being developed, a request for proposal written, bids from contractors sourced and evaluated, and a winner chosen. The contractor, once hired, brings a team to develop the software based on the RFP, and the staffing levels (counting both internal and contracting staff) shoot up. There is a development period, followed by a short period of “user acceptance testing,” and then the project falls into “operations and maintenance,” which is generally a different “color of money” than the development funds.



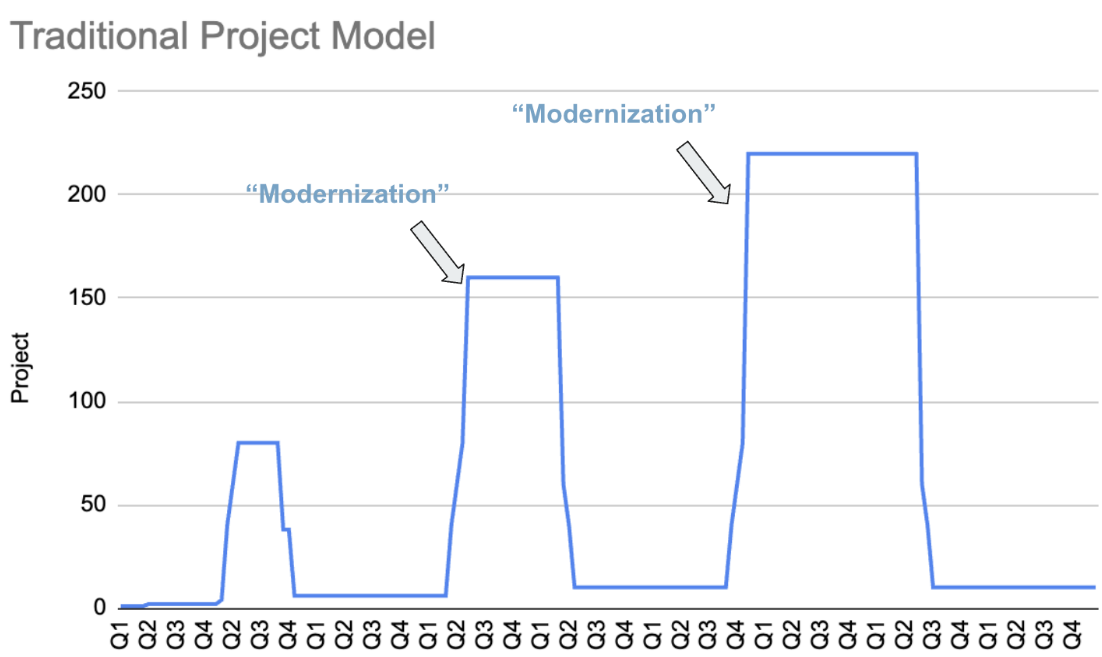
Contrast this with a typical “product” model, in which, instead of a requirements gathering phase up front, a small team, often but not always internal to government, conducts what are called discovery sprints to better understand the problems the software is supposed to address. If some parts of the proposed solution are riskier than others (for instance, it’s not clear whether a data integration will work well), they find ways to test those problems first, before an entire software solution has been built. They may develop prototypes to help question their assumptions, and they engage with users from the beginning. Product teams almost always leverage contractors, but the contractors are there to complement a core internal team which holds the product vision and provides clear direction to vendors. Staff is added slowly over time as the team learns what they need, but doesn't dramatically ramp down once a first or even second version is shipped. As my colleague Dave Guarino quips, "Google didn't lay everyone off after they put up search." Indeed, they invested more.



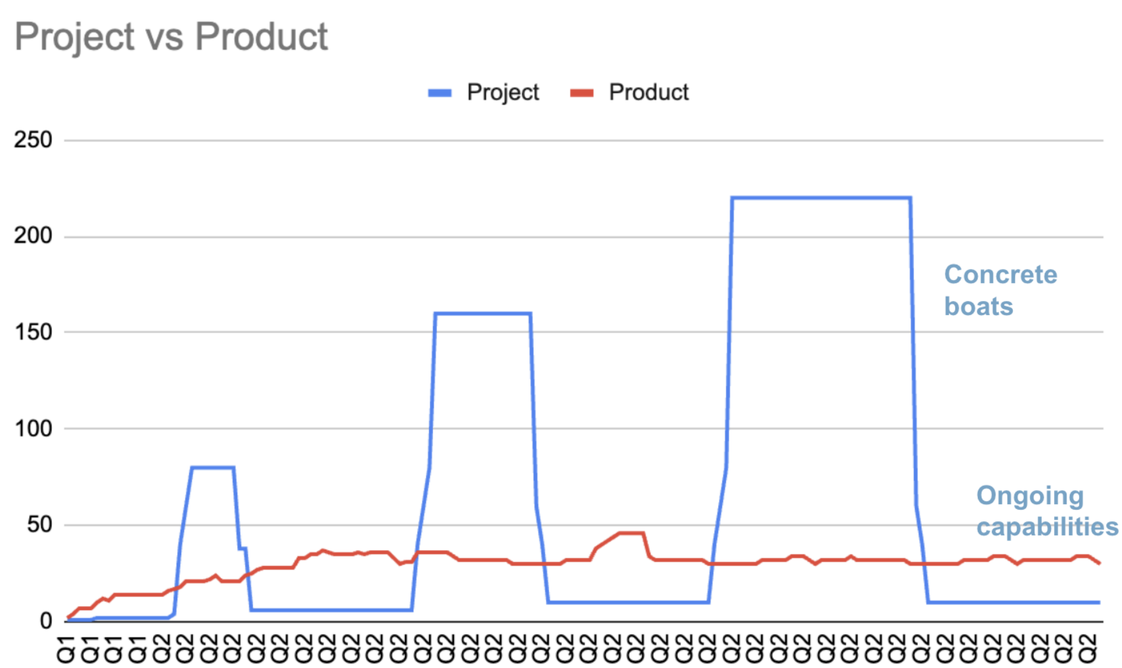
At this timescale, there seems to be an obvious reason to prefer the project model: the minimal ongoing expense. But as anyone following government technology appropriations knows, this is not the right timescale to look at. What happens next on the project line is one or more of the following scenarios: the software doesn't work well for its users, and funds are sought to fix its defects; it quickly becomes outdated, either by changes in the technical environment, the policy environment, or other external factors, and funds are sought for modernization; or new needs have emerged that the existing software doesn't address. Thus, the actual project model line looks more like this in the medium term:



And then in the longer term, as modernizations fail, needs escalate, and even more money is allocated, like this:



Here is where the slow and steady product line starts to look more attractive, on a purely cost basis, though cost is far from the only reason to prefer it. Having a consistent team over time may look like an unwanted ongoing expense, if we assume that development work at some point "is done," but that is not the case. ("Software is never done" was one of the precepts of the Software Acquisition Practices report I contributed to for the Defense Innovation Board under President Trump[4]). The product model is not only less expensive in the long run, it results in working software that rarely needs "modernization" of the kind you've become used to hearing about because it's constantly being updated and improved. The biggest difference between the project and product models is that the steady investment over time delivers effective service to the American people. Periodic investment in "projects" is how we get backlogs and confused and frustrated constituents.



[4] https://innovation.defense.gov/software/

A summary of the differences in these models is below:

Project model	Product model
Episodic large investments	Ongoing moderate investment
Heavy reliance on procurement, oversight, and IV&V	Requires internal product ownership and management
Abdication of responsibility to vendors	Partnership w/vendors
Vendor lock-in	Low switching costs, smaller contracts
Acquiring static software	Acquiring ongoing capabilities
Constant loss of knowledge	Constantly growing understanding
Customers consulted at end	Customers integral at all times
Built in silos	The walls come down
Subject to the "100% trap"	85% to start costs 10% of the price
High rates of failure and frustration	Actual working software

As much as it's tempting to hold agencies accountable for their addiction to the project model, this is not something they can fix on their own. Congress would need to enable ongoing funding streams (in addition to procurement changes previously discussed) in order to see agencies develop in the product model. Given all that we know about how fast AI moves, the risks it carries, and the benefits it could bring, Congress should work with OMB and agencies to change the laws, regulations, processes, and practices that impede agencies from operating in the product model.

Workforce

How will we get the AI workforce we need? OpenAI famously recruits talent with $1M signing bonuses. Government can't compete on compensation, and it likely never will in this domain. But it competes remarkably well these days by selling the mission. For many in tech, the mission is irresistible. Organizations like the Tech Talent Project, which place digital professionals in roles in federal and state agencies, now have backlogs of tech leaders eager to serve the American public. For some, it was the pandemic's brutal reminder of how much government matters. For others, it is threats to our country's standing in the world that they want to counter. Whatever the reason, we now have people willing, and our greatest leverage will be in fixing the systems needed to actually hire them.

You would think that when we have proven tech talent ready to serve, we would jump to bring them on quickly. In fact, that backlog of tech leaders eager to join is largely languishing in hiring processes that can easily take nine months or longer. This could change, but it will require taking seriously the defects in our hiring practices. It's not just speed, but how we hire. Today, 90% of competitive, open-to-the-public job announcements across the federal government rely solely on a resume review by an HR generalist and an applicant's self-assessment of their skills.[5] In other words, we have essentially one way to determine if candidates are qualified for the vast majority of positions — we ask them to rate themselves. Hiring managers often receive

[5]GSA's Hiring Assessment and Selection Outcome Dashboard https://d2d.gsa.gov/report/hiring-assessment-and-selection-outcome-dashboard

from HR staff a slate that contains no qualified candidates, which is why half of all hiring actions fail. They simply reject these slates and start over, adding even more months to what is already an unacceptable timeframe. Meanwhile, they miss qualified candidates whose resumes didn't make the cut because they didn't know the absurd games applicants must play to get placed on the HR hiring slates, like copying and pasting the qualifications noted in the job description directly into their resume, and rating themselves "master" at every single competency listed in the assessment. We are losing too many willing digital professionals, not because of lower pay, but because of arcane, cumbersome processes. Lack of flexibilities like remote work makes the problem even worse.

The Office of Personnel Management (OPM) and the White House have stated their intentions to hire the AI tech talent needed, but this is a case where strengthening the workforce is also a matter of reducing burdens. OPM's recent memo, for instance, will grant direct hire authority for several AI-related job classifications. That will remove a bit of the red tape agencies need to bring on experts. But that direct hire authority does not allow for the use of pooled hiring across agencies, despite the fact that pooled hiring has gotten us many excellent data scientists and other tech roles much more quickly. Agencies will have to run a separate hiring action for each open position, which will take enormous amounts of time and paperwork, even with the direct hire authority. Congress should ask OPM what authorities they need in order to change this, and what resources they need to scale programs like the highly successful Subject Matter Expert Qualifying Assessment (SME-QA) program. Then ask what is the next obstacle they need removed. I don't presume to know everything that is needed, only that they operate in a highly constrained environment, no longer fit to the purpose it must serve.

For those who despair of our ability to compete for talent, it's important to remember that the people OpenAI and others are hiring at such sky high salaries are typically those who know how to *develop and train* models. Government's primary need is not for that very specialized talent pool. It is for people who know how to *use* these models. Though I am a fan of the notion of government creating its own models, that will be the extremely rare exception. The commercial and open source communities will provide models government can adopt. The expertise needed to take advantage of AI software developed by others is at far less of a premium than that of the talent pool getting the $1M signing bonuses, and it is even more critical to successful adoption. The kinds of technologists that USDS, 18F, and federal agencies have been hiring and continue to hire – service designers, product managers, data engineers – can do this work, even if they are not technically experts in AI (though some are). We just need to hire them at a much greater scale.

Greater competence and capacity are also important because we need people who use AI, when appropriate, to solve real problems. There is the very real risk that agencies, especially those that lack sufficient basic digital expertise, will buy AI tools in ways that are compliant with all the new guidance, but that fundamentally lack an understanding of the problem they are trying to solve. We've seen this many times before in government and elsewhere, especially with blockchain technologies – a rush to sprinkle "advanced tech fairy dust" on a tech portfolio without a clear purpose or a clear match between the need and the solution. These thoughtless implementations will harm the public, give AI in customer service a bad name, and understandably strengthen the calls to slow down. The more uses of AI for AI's sake, the more we risk stifling what could be a welcome advancement if done thoughtfully.

AI can't be done thoughtfully without the right workforce. And we can't legislate our way to the right workforce, though removing previous legislative mandates may help. Congress will need to encourage and enable OPM to build the human resources system we need to meet this moment.

An Enablement Approach

It can be difficult to legislate competence in digital or any other domain. A large part of what makes us bad at customer experience in a digital age is that we have created a system in which the careers of government staff depend more on compliance with process than on achieving the desired outcomes. More rules usually exacerbate this effect, leading, ironically, to worse outcomes. Even legislation that doesn't add rules, but simply directs the executive branch to make studies or plans can lead to more unhelpful rigidity.

In my book *Recoding America* I tell a story of a team unable to ship the software for the new GPS satellites because they've been told that a certain component, one that breaks the software, is required by law. Many people up and down the hierarchy literally believe that Congress has mandated this component. Because of this belief, no one can get approval to take it out, even though the software has gone years over schedule and billions of dollars over budget, and would finally work if this component were removed. (It never was.)

Congress, of course, had not mandated that this specific component be used in this specific software project. In the 1996 Clinger Cohen Act, Congress had mandated that OMB provide high level guidance around interoperability in software, and this component was used to illustrate how interoperability *might* be achieved. As that high level guidance was translated into ever more concrete and binding policies at lower levels of within government, risk aversion caused it to go from an illustration to a recommendation to a binding requirement, in this case dooming the project. Even when legislation is written with sufficient leeway to allow implementers to use their judgment, it runs the risk of causing the sort of calcification that leads to bad outcomes. We must be careful what we legislate lest it have negative unintended consequences.

The goal, therefore, must be Congressional action that reduces the risk aversion of the bureaucracy. Simplification of accumulated policy cruft as described above (with help from AI) falls into this category. Careful use of oversight, including to lift up successes as often as we question failures, counts as well. Getting government agencies the people they need, focused on the right work, and reducing the burden on each of them, can be profoundly transformational.

Conclusion

As we enter the AI era, we are forced to finally grapple with the lessons of the Internet era. Chief among those lessons is how much lack of digital capacity in government has hurt the American people. Fortunately, we already know much of what we need to do to face this challenge, because it is largely the same work we have needed to do for the past two decades. AI is still software, just software that intensifies and speeds up the need for change that we've observed to date. Its arrival is our wakeup call to do what we should have already done, but it is also a gift that will help us do this work.