

December 5, 2023

**Response to the Request for Comments on OMB's
Advancing Governance, Innovation, and Risk Management
for Agency Use of Artificial Intelligence Draft Memorandum**

Data & Society is an independent, nonprofit research institute studying the social implications of data-centric technologies, automation, and artificial intelligence. Through empirical research and policy and media engagement, our work illuminates the values and decisions that drive these systems and helps shape futures grounded in equity and human dignity.

OMB's draft guidance takes important steps toward AI accountability. As detailed in response to the questions below, the OMB guidance rightly identifies many AI systems that have caused harm to the public and appropriately requires federal agencies to institute key practices to protect the public against further such harms.

Critically, the OMB guidance starts with the presumption that certain uses of AI impact people's rights. The procedure to determine whether an AI use case is rights-impacting, described in the draft memo, includes a predetermined list of systems where harms have previously been demonstrated and which are presumed to require the described risk mitigation practices. Employing this presumption based on the predetermined lists is a useful mechanism that has the potential to allow agencies to guard against harms without being required to conduct extensive and potentially time-consuming risk assessments in cases where AI system harms are already publicly known. Essentially, OMB has already conducted the risk assessment on their behalf.

The memo mandates a floor of minimum practices that an agency must meet in order to use an AI system. The required minimum practices for rights-impacting AI systems include sociotechnical approaches to AI accountability, including impact assessments and public consultation. Such approaches, appropriately used in combination, represent the best known standards for AI accountability across a wide variety of AI systems. The OMB memo requires that if minimum practices are not met, or if testing required by these practices reveals concerns, agencies cannot use the system in question. This prohibition on unrestricted use is key to future effectiveness of this guidance.

We applaud these steps and encourage OMB to keep these important measures in their final guidance. The procedures described by the draft memorandum are both achievable and necessary to protect the public from the demonstrated harms of rights-impacting AI systems. We encourage OMB to preserve these needed protections in its final guidance, and particularly to maintain the requirement that federal agencies may not use systems that fail to meet the identified minimum practices.

In response to OMB's specific questions, we provide the following additional thoughts.

Table of Contents

3. How can OMB best advance responsible AI innovation?2

5. Are there use cases for presumed safety-impacting and rights-impacting AI (Section 5 (b)) that should be included, removed, or revised? If so, why?3

 AI systems that should be presumed to be rights-impacting3

 Clarification to the definition of AI12

 Waiver process.....12

6. Do the minimum practices identified for safety-impacting and rights-impacting AI set an appropriate baseline that is applicable across all agencies and all such uses of AI? How can the minimum practices be improved, recognizing that agencies will need to apply context-specific risk mitigations in addition to what is listed?13

 Consult and incorporate feedback from affected groups (Section 5.c.v.B.).....13

 Missing practices14

 Data privacy.....14

 Explanation15

7. What types of materials or resources would be most valuable to help agencies, as appropriate, incorporate the requirements and recommendations of this memorandum into relevant contracts?15

 Concern that the values language in the draft memorandum does not center equity16

 Contract requirements that include the minimum practices17

 Detailed reporting and the AI use case inventory.....17

 Federal grants must also require the minimum practices17

8. What kind of information should be made public about agencies' use of AI in their annual use case inventory?18

 Including all AI systems in the inventory18

 Summary reporting of excluded AI systems19

 Detailed reporting of testing and risk as part of the AI use case inventory19

 Complete an AI impact assessment (Section 5.c.iv.A).....19

 Take steps to ensure that the AI will advance equity, dignity, and fairness (Section 5.c.v.A)20

 Consult and incorporate feedback from affected groups (Section 5.c.v.B).....21

3. How can OMB best advance responsible AI innovation?

Key to the private sector’s successful use of responsible AI innovation is the federal government’s oversight of that use in ways that support safety and rights. As part of the National AI Talent Surge, enforcement agencies such as the Department of Justice Civil Rights Division, the Federal Trade Commission, the Equal Employment Opportunity Commission, and the Consumer Financial Protection

Bureau must receive significantly more staffing. Such agencies are often overlooked when considering technical talent, but should be provided staffing to build AI auditing systems and conduct assessments.

5. Are there use cases for presumed safety-impacting and rights-impacting AI (Section 5 (b)) that should be included, removed, or revised? If so, why?

AI systems that should be presumed to be rights-impacting

Research and reporting have demonstrated harms from the AI systems that are already included in OMB's list of systems presumed to be rights-impacting. In support of keeping *all* such systems on this list and requiring they meet a minimum bar of protective practices, we include references to the documented harms of these systems below. Additionally, we suggest below specific systems that should be added to the list based on similarly documented harms. We note that the preface to the list of systems is appropriate and should not be substantially modified — it is important that the list not only includes AI systems that directly control outcomes but those that influence them, since a human is often in the loop for rights-impacting decisions. The preface states that the below AI systems should be presumed to be rights-impacting AI "if it is used to control or meaningfully influence the outcomes of any of the following activities or decisions":

Purposes That Are Presumed to Be Rights-Impacting.

- A. Decisions to block, remove, hide, or limit the reach of protected speech;¹
- B. Law enforcement or surveillance-related risk assessments about individuals,² criminal recidivism prediction,³ offender prediction,⁴ predicting perpetrators' identities,⁵ victim prediction,⁶ crime

¹ Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018, December). Measuring and mitigating unintended bias in text classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 67-73); Thompson, A. (2017, October 25). Google's Sentiment Analyzer Thinks Being Gay Is Bad. *Vice*. <https://www.vice.com/en/article/j5jmi8/google-artificial-intelligence-bias> ; Jessica Guynn. Facebook while black: Users call it getting 'Zucked,' say talking about racism is censored as hate speech. USA Today. Apr. 24, 2019. <https://www.usatoday.com/story/news/2019/04/24/facebook-while-black-zucked-users-say-they-get-blocked-racism-discussion/2859593002/>

² Brayne, S. (2020). Predict and Surveil: Data, discretion, and the future of policing. Oxford University Press, USA.; Joh, E. E. (2016). The new surveillance discretion: Automated suspicion, big data, and policing. *Harvard Law and Policy Review*, 10, 15.

³ Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> ; Dressel, J., & Farid, H. (2018, January 17). The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1). <https://doi.org/10.1126/sciadv.aao5580>

⁴ Stroud, M. (2021, May 24). Heat Listed. *The Verge*. <https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list>

⁵ Brayne, S. (2020). Predict and Surveil: Data, discretion, and the future of policing. Oxford University Press, USA.

⁶ Lipari, J. L. (2020, January 23). Advisory Concerning the Chicago Police Department's Predictive Risk Models. *Report of the Public Safety Section of The Office of Inspector General, OIG FILE #18-0106*. <https://igchicago.org/wp-content/uploads/2020/01/OIG-Advisory-Concerning-CPDs-Predictive-Risk-Models-.pdf> ; Mick Dumke and Frank Main. A look inside the watch list Chicago police fought to keep secret. *The Chicago Sun Times*. May 18, 2017. <https://chicago.suntimes.com/2017/5/18/18386116/a-look-inside-the-watch-list-chicago-police-fought-to-keep-secret>

forecasting,⁷ license plate readers,⁸ iris matching,⁹ facial matching,¹⁰ facial sketching,¹¹ genetic facial reconstruction,¹² social media monitoring,¹³ prison monitoring,¹⁴ forensic analysis,¹⁵ forensic genetics,¹⁶ the conduct of cyber intrusions,¹⁷ physical location-monitoring devices,¹⁸ or decisions related to sentencing, parole, supervised release, probation, bail, pretrial release, or pretrial detention;¹⁹

⁷ Lum, K., & Isaac, W. (2016). To predict and serve?. *Significance*, 13(5), 14-19.; Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018, January). Runaway feedback loops in predictive policing. In *Conference on fairness, accountability and transparency* (pp. 160-171). PMLR.; Sankin, A., Mehrotra, D., Mattu, S. and Gilbertson, A. (2021, December 2). Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them. *The Markup*. <https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them>; Sankin, A. and Mattu, S. (2023, October 2). Predictive Policing Software Terrible At Predicting Crimes. *The Markup*. <https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes>

⁸ Stein, N. (2023, January). Automated License Plate Readers: Legal and Policy Evaluation. *University of Michigan Ford School of Public Policy*. https://stpp.fordschool.umich.edu/sites/stpp/files/2023-02/ALPR%20Memo%20Final%20Jan%202023_0.pdf

⁹ Electronic Frontier Foundation. Iris Recognition. <https://www.eff.org/pages/iris-recognition>

¹⁰ Hill, K. (2020, June 24). Wrongfully Accused by an Algorithm. *The New York Times*.

<https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>; Hill, K. (2023, August 6). Eight Months Pregnant and Arrested after False Facial Recognition Match. *The New York Times*.

<https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html>; Johnson, K. (2023, February 28). Face Recognition Software Led to His Arrest. It Was Dead Wrong. *WIRED*.

<https://www.wired.com/story/face-recognition-software-led-to-his-arrest-it-was-dead-wrong/>

¹¹ Chloe Xiang. (2023, February 7). Developers Created AI to Generate Police Sketches. Experts Are Horrified. *Motherboard*. <https://www.vice.com/en/article/qjk745/ai-police-sketches>

¹² Southall, A. (2017, October 19). Using DNA to Sketch What Victims Look Like; Some Call It Science Fiction. *The New York Times*. <https://www.nytimes.com/2017/10/19/nyregion/dna-phenotyping-new-york-police.html>;

Xiang, C. (2022, October 11). Police Are Using DNA to Generate 3D Images of Suspects They've Never Seen. *Motherboard*. <https://www.vice.com/en/article/pkgma8/police-are-using-dna-to-generate-3d-images-of-suspects-theyve-never-seen>

¹³ Levinson-Waldman, R. (2018). Government Access to and Manipulation of Social Media: Legal and Policy Challenges. *Howard Law Journal*, 61, 3. p. 523-562.

¹⁴ Bender, E. M. and Tatman, R. (2021, October 9). AI surveillance in prisons is a terrible idea, both technologically and ethically. *Geek Wire*. <https://www.geekwire.com/2021/guest-post-ai-surveillance-prisons-terrible-idea-technologically-ethically/>;

Asher-Schapiro, A. and Sherfinski, D. (2021, November 16). 'Scary and chilling': AI surveillance takes U.S. prisons by storm. *Reuters*. <https://www.reuters.com/article/usa-prisons-surveillance-idUSKBN2I01H0/>

¹⁵ Wexler, R. (2018, May). Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System. *Stanford Law Review*, 70, 5. <https://www.stanfordlawreview.org/print/article/life-liberty-and-trade-secrets/>

¹⁶ Ibid.

¹⁷ Ibid.

¹⁸ Glaser, A. (2021, July 5). Incarcerated at home: The rise of ankle monitors and house arrest during the pandemic. *NBC News*. <https://www.nbcnews.com/tech/tech-news/incarcerated-home-rise-ankle-monitors-house-arrest-during-pandemic-n1273008>

¹⁹ Cohen, T. H., Lowenkamp, C. T., and Hicks, W. E. (2018, September). Revalidating the Federal Pretrial Risk Assessment Instrument (PTRA): A Research Summary. *Federal Probation*, 82, 2.

https://www.uscourts.gov/sites/default/files/82_2_3_0.pdf; Coalition letter. (2020, April 3). RE: The use of the PATTERN risk assessment in prioritizing release in response to the COVID-19 pandemic. https://www.upturn.org/static/files/Final_Letter_on_PATTERN_in_Response_to_AG_Barr_Memo_on_4_26-4_3_2020.pdf

- a. Suggested change: ensure that the closing clause includes "decisions *or risk assessments* related to ..." to ensure that, for example, pretrial risk assessments²⁰ are included in this list.
- C. Deciding immigration, asylum, or detention status;²¹ providing risk assessments about individuals who intend to travel to, or have already entered, the U.S. or its territories;²² determining border access or access to Federal immigration related services through biometrics (e.g., facial matching) or other means (e.g., monitoring of social media or protected online speech);²³ translating official communication to an individual in an immigration, asylum, detention, or border context;²⁴ or immigration, asylum, or detention-related physical location monitoring devices.²⁵
 - a. Suggested change: in the first clause "deciding" should be changed to "deciding or providing risk assessments related to" to ensure that detention risk assessments²⁶ are included.

²⁰ The Leadership Conference for Civil and Human Rights. The use of pre-trial "risk assessment" instruments: a shared statement of civil rights concerns. <http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf>; Carrie Johnson. Flaws plague a tool meant to help low-risk federal prisoners win early release. NPR. Jan. 26, 2022. <https://www.npr.org/2022/01/26/1075509175/flaws-plague-a-tool-meant-to-help-low-risk-federal-prisoners-win-early-release>; Carrie Johnson. Justice Department works to curb racial bias in deciding who's released from prison. NPR. Apr. 19, 2022. <https://www.npr.org/2022/04/19/1093538706/justice-department-works-to-curb-racial-bias-in-deciding-whos-released-from-pris>; National Institute of Justice. 2021 Review and Revalidation of the First Step Act Risk Assessment Tool. National Institute of Justice NCJ 303859. Dec., 2021. <https://www.ojp.gov/pdffiles1/nij/303859.pdf>

²¹ Petra Molnar and Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System," (University of Toronto International Human Rights Program and the Citizen Lab, Munk School of Global Affairs, 2018) <https://citizenlab.ca/2018/09/bots-at-the-gate-human-rights-analysis-automated-decision-making-in-canadas-immigration-refugee-system/>; Koulisch, R., & Evans, K. (2021). Punishing with Impunity: The Legacy of Risk Classification Assessment in Immigration Detention. *Geo. Immigr. LJ*, 36, 1. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=6813&context=faculty_scholarship; Evans, K., & Koulisch, R. (2020). Manipulating risk: immigration detention through automation. *Lewis & Clark L. Rev.*, 24, 789. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=6692&context=faculty_scholarship

²² Levinson-Waldman, R. and Guillermo Gutiérrez, J. (2023, October 19). Overdue Scrutiny for Watch Listing and Risk Prediction: Reining In Civil Liberties Abuses and Assessing Efficacy. *Brennan Center for Justice Report*.

²³ Vavoula, N. (2021). Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism, *European Journal of Migration and Law*, 23(4), 457-484. doi: <https://doi.org/10.1163/15718166-12340114>; UN Special Rapporteur on Contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Report to the 75th session of the UN General Assembly. November 2020 <https://undocs.org/A/75/590>

²⁴ Nicholas, G., & Bhatia, A. (2023). Lost in Translation: Large Language Models in Non-English Content Analysis. *Center for Democracy and Technology Research Report*. <https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf>

²⁵ Aguilera, J. (2022, April 18). U.S. Officials Deploy Technology to Track More Than 200,000 Immigrants, Triggering a New Privacy Lawsuit. *Time Magazine*. <https://time.com/6167467/immigrant-tracking-ice-technology-data/>; Bhuiyan, J. (2022, March 8). 'Constantly afraid': immigrants on life under the US government's eye. *The Guardian*. <https://www.theguardian.com/us-news/2022/mar/08/us-immigrants-isap-ice-bi-ankle-monitor>

²⁶ Koulisch, R., & Evans, K. (2021). Punishing with Impunity: The Legacy of Risk Classification Assessment in Immigration Detention. *Geo. Immigr. LJ*, 36, 1. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=6813&context=faculty_scholarship; Evans, K., &

- b. Suggested additions: surveillance systems used to support routine immigration enforcement²⁷
- D. Detecting or measuring emotions, thought, or deception in humans,²⁸
- E. In education, detecting student cheating or plagiarism,²⁹ influencing admissions processes,³⁰ monitoring students online or in virtual-reality,³¹ projecting student progress or outcomes,³² recommending disciplinary interventions,³³ determining access to educational resources or

Koulish, R. (2020). Manipulating risk: immigration detention through automation. *Lewis & Clark L. Rev.*, 24, 789. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=6692&context=faculty_scholarship

²⁷ Dwyer, M. P. and Levinson-Waldman, R. (2023, June 29). A Realignment for Homeland Security Investigations. *Brennan Center for Justice Report*. <https://www.brennancenter.org/our-work/research-reports/realignment-homeland-security-investigations>

²⁸ Stark, L., & Hoey, J. (2021, March). The ethics of emotion in artificial intelligence systems. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 782-793). <https://dl.acm.org/doi/pdf/10.1145/3442188.3445939> ; Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press. Chapter: Affect.; Engler, A. (2021, August 4). Why President Biden should ban affective computing in federal law enforcement. *Brookings*. <https://www.brookings.edu/articles/why-president-biden-should-ban-affective-computing-in-federal-law-enforcement/>

²⁹ Mathewson, T. G. (2023, August 14). AI Detection Tools Falsely Accuse International Students of Cheating: Stanford study found AI detectors are biased against non-native English speakers. *The Markup*. <https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating> ; Hill, K. (2022, May 27). Accused of Cheating by an Algorithm, and a Professor She Had Never Met. *The New York Times*. <https://www.nytimes.com/2022/05/27/technology/college-students-cheating-software-honorlock.html> ; National Disabled Law Students Association. Report on Concerns Regarding Online Administration of Bar Exams. Jul. 29, 2020. https://ndlsa.org/wp-content/uploads/2020/08/NDLSA_Online-Exam-Concerns-Report1.pdf ; Lydia X. Z. Brown. How Automated Test Proctoring Software Discriminates Against Disabled Students. Center for Democracy and Technology. Nov. 16, 2020. <https://cdt.org/insights/how-automated-test-proctoring-software-discriminates-against-disabled-students/>

³⁰ Engler, A. (2021, September 14). Enrollment algorithms are contributing to the crises of higher education. *Brookings*. <https://www.brookings.edu/articles/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/>

³¹ Laird, E., Grant-Chapman, H., Venzke, C., and Quay-de la Vallee, H. (2022, August 3). Report – Hidden Harms: The Misleading Promise of Monitoring Students Online. *Center for Democracy and Technology*. <https://cdt.org/insights/report-hidden-harms-the-misleading-promise-of-monitoring-students-online/>

³² Feathers, T. (2023, April 27). False Alarm: How Wisconsin Uses Race and Income to Label Students “High Risk”. *The Markup*. <https://themarkup.org/machine-learning/2023/04/27/false-alarm-how-wisconsin-uses-race-and-income-to-label-students-high-risk>

³³ Quay-de la Vallee, H. & Duarte, N. (2019, August 19). Algorithmic Systems in Education: Incorporating Equity and Fairness When Using Student Data. *Center for Democracy and Technology*. <https://cdt.org/insights/algorithmic-systems-in-education-incorporating-equity-and-fairness-when-using-student-data/>.

programs,³⁴ determining eligibility for student aid,³⁵ or facilitating surveillance (whether online or in-person);³⁶

- a. Suggested additions: AI used as part of student advising,³⁷ or facial recognition or other biometrics used in the context of education.³⁸
- F. Tenant screening or controls,³⁹ home valuation,⁴⁰ mortgage underwriting,⁴¹ or determining access to or terms of home insurance;⁴²
 - a. Suggested additions: housing advertising,⁴³ facial recognition or other biometrics when used in the context of public housing.⁴⁴

³⁴ Lecher, C. and Varner, M. NYC’s School Algorithms Cement Segregation. This Data Shows How. *The Markup*. <https://themarkup.org/machine-learning/2021/05/26/nycs-school-algorithms-cement-segregation-this-data-shows-how>

³⁵ Engler, A. (2021, September 14). Enrollment algorithms are contributing to the crises of higher education. *Brookings*. <https://www.brookings.edu/articles/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/>; Student Borrower Protection Center. Educational Redlining. Student Borrower Protection Center Report. Feb. 2020. <https://protectborrowers.org/wp-content/uploads/2020/02/Education-Redlining-Report.pdf>

³⁶ Laird, E., Grant-Chapman, H., Venzke, C., and Quay-de la Vallee, H. (2022, August 3). Report – Hidden Harms: The Misleading Promise of Monitoring Students Online. *Center for Democracy and Technology*. <https://cdt.org/insights/report-hidden-harms-the-misleading-promise-of-monitoring-students-online/>; Quay-de la Vallee, H. (2022). The Chilling Effect of Student Monitoring: Disproportionate Impacts and Mental Health Risks. *Center for Democracy and Technology*. <https://cdt.org/insights/the-chilling-effect-of-student-monitoring-disproportionate-impacts-and-mental-health-risks/>; Gillum, J. and Kao, J. (2019, June 25). Aggression Detectors: The Unproven, Invasive Surveillance Technology Schools Are Using to Monitor Students. *ProPublica*. <https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/>

³⁷ Blume, H. (2023, August 7). AI Chatbot ‘Ed’ Will Be L.A. Unified’s Newest Student Adviser, Superintendent Says. *Education Week*. <https://www.edweek.org/leadership/ai-chatbot-ed-will-be-l-a-unifieds-newest-student-adviser-superintendent-says/2023/08>; Feathers, T. (2021, March 2). Major Universities Are Using Race as a “High Impact Predictor” of Student Success. *The Markup*. <https://themarkup.org/machine-learning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success>

³⁸ ACLU of New York. What You Need to Know About New York’s Temporary Ban on Facial Recognition in Schools. Accessed May 2, 2022. <https://www.nyclu.org/en/publications/what-you-need-know-about-new-yorks-temporary-ban-facial-recognition-schools>

³⁹ Kirchner, L. and Goldstein, M. (2020, May 28). Access Denied: Faulty Automated Background Checks Freeze Out Renters. *The Markup and The New York Times*. <https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters>

⁴⁰ Sara Safransky, “Geographies of Algorithmic Violence: Redlining the Smart City,” *International Journal of Urban and Regional Research*, Nov. 24, 2019, 9, available at <https://onlinelibrary.wiley.com/doi/full/10.1111/1468-2427.12833>.

⁴¹ Emmanuel Martinez, Lauren Kirchner, (2021, Aug. 25) “The Secret Bias Hidden in Mortgage-Approval Algorithms,” *The Markup*, available at <https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms>.

⁴² Ronda Lee (2022, Nov. 1). AI can perpetuate racial bias in insurance underwriting. *Yahoo! Money*. <https://money.yahoo.com/ai-perpetuates-bias-insurance-132122338.html?guccounter=1>.

⁴³ Department of Justice. (2022, June 21). Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising. <https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known>

⁴⁴ Fadulu, L. (2019, September 24). Facial Recognition Technology in Public Housing Prompts Backlash. *The New York Times*. <https://www.nytimes.com/2019/09/24/us/politics/facial-recognition-technology-housing.html>

- G. Determining the terms and conditions of employment,⁴⁵ including pre-employment screening,⁴⁶ pay or promotion,⁴⁷ performance management,⁴⁸ hiring or termination,⁴⁹ time-on-task tracking,⁵⁰ virtual or augmented reality workplace training programs,⁵¹ or electronic workplace surveillance and management systems;⁵²
- a. Suggested additions: algorithmic disciplinary actions based on third party ratings,⁵³ automated scheduling,⁵⁴ or worker classification.⁵⁵

⁴⁵ Nguyen, A. (2021, May 19). The Constant Boss. *Data & Society*. <https://datasociety.net/library/the-constant-boss/>; Mateescu, A. (2023, November 8). Challenging Worker Datafication. *Data & Society*. <https://datasociety.net/library/challenging-worker-datafication/>.

⁴⁶ Jeffrey Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Oct. 10, 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

⁴⁷ Dubal, V. (2023, January 19). On Algorithmic Wage Discrimination. UC San Francisco Research Paper No. Forthcoming. <https://ssrn.com/abstract=4331080>; Lauren Kaori Gurley. Amazon's AI Cameras Are Punishing Drivers for Mistakes They Didn't Make. Motherboard. Sep. 20, 2021. <https://www.vice.com/en/article/88npjv/amazons-ai-cameras-are-punishing-drivers-for-mistakes-they-didnt-make>

⁴⁸ Mateescu, M. & Nguyen, A. (2019, Feb.). Explainer: Algorithmic Management in the Workplace. *Data & Society*. <https://datasociety.net/library/explainer-algorithmic-management-in-the-workplace/>.

⁴⁹ Miranda Bogen. All the Ways Hiring Algorithms Can Introduce Bias, *Harvard Business Review* (May 6, 2019). <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias>; Spencer Soper. Fired by Bot at Amazon: "It's You Against the Machine". Bloomberg, Jun. 28, 2021. <https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out>

⁵⁰ Colin Lecher, How Amazon automatically tracks and fires warehouse workers for 'productivity.' *The Verge* (April 5, 2019), <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations>

⁵¹ boyd, d. (2014). Is the Oculus Rift sexist? *Quartz*. <https://qz.com/192874/is-the-oculus-rift-designed-to-be-sexist> ; MacArthur, C., Grinberg, A., Harley, D., & Hancock, M. (2021, May). You're making me sick: A systematic review of how virtual reality research considers gender & cybersickness. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-15). <https://dl.acm.org/doi/abs/10.1145/3411764.3445701>

⁵² Scherer, M., & Brown, L. X. (2021). Warning: Bossware may be hazardous to your health. Center for Democracy and Technology. <https://cdt.org/wp-content/uploads/2021/07/2021-07-29-Warning-Bossware-May-Be-Hazardous-To-Your-Health-Final.pdf> ; Human Impact Partners and WWRC. The Public Health Crisis Hidden in Amazon Warehouses. HIP and WWRC report. Jan. 2021. <https://humanimpact.org/wp-content/uploads/2021/01/The-Public-Health-Crisis-Hidden-In-Amazon-Warehouses-HIP-WWRC-01-21.pdf>

⁵³ Nguyen, A. & Zelickson, E. (2022, Oct. 12). At the Digital Doorstep. *Data & Society*. <https://datasociety.net/library/at-the-digital-doorstep/>.

⁵⁴ Kaye Loggins. Here's What Happens When an Algorithm Determines Your Work Schedule. *Vice* (Feb. 24, 2020). <https://www.vice.com/en/article/g5xwby/heres-what-happens-when-an-algorithm-determines-your-work-schedule>.

⁵⁵ Mateescu, A. & Nguyen, A. (2019, Feb.). Explainer: Algorithmic Management in the Workplace. *Data & Society*. <https://datasociety.net/library/explainer-algorithmic-management-in-the-workplace/>.

- H. Decisions regarding medical devices,⁵⁶ medical diagnostic tools,⁵⁷ clinical diagnosis and determination of treatment,⁵⁸ medical or insurance health-risk assessments,⁵⁹ drug-addiction risk assessments and associated access systems,⁶⁰ suicide or other violence risk assessment,⁶¹ mental-health status detection or prevention,⁶² systems that flag patients for interventions,⁶³ public insurance care-allocation systems,⁶⁴ or health-insurance cost and underwriting processes;⁶⁵
- a. Suggested additions: AI systems used to summarize doctor's notes and/or analyze health records⁶⁶

⁵⁶ Wu, E., Wu, K., Daneshjou, R. *et al.* How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 27, 582–584 (2021). <https://doi.org/10.1038/s41591-021-01312-x>.

⁵⁷ DeSario, G. D. *et al.* (2023). Using AI to Detect Pain Through Facial Expressions: A Review.

Bioengineering, 10, 5. p548. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10215219/>;

Note the retraction of this paper: Hosseini, M. *et al.* (2021). Retraction: Deep Learning for Autism Diagnosis and Facial Analysis in Children. *Frontiers in Computational Neuroscience*, 15.

<https://www.frontiersin.org/articles/10.3389/fncom.2021.789998/full> ;

⁵⁸ Ferryman, K. and Pitcan, M. (2018, February 26). Fairness in Precision Medicine. *Data & Society*.

<https://datasociety.net/library/fairness-in-precision-medicine/> ; Andrew Wong *et al.* External validation of a widely

implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med.* 2021; 181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626 ; Darshali A. Vyas *et al.*, Hidden in Plain Sight – Reconsidering the

Use of Race Correction in Clinical Algorithms, 383 *N. Engl. J. Med.* 874, 876-78 (Aug. 27, 2020),

<https://www.nejm.org/doi/full/10.1056/NEJMms2004740>

⁵⁹ Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

<https://www.science.org/doi/full/10.1126/science.aax2342> ; Angela Chen. Why the Future of Life Insurance May

Depend on Your Online Presence. *The Verge*. Feb. 7, 2019. <https://www.theverge.com/2019/2/7/18211890/social-media-life-insurance-new-york-algorithms-big-data-discrimination-online-records>

⁶⁰ Szalavitz, M. (2021, August 11). The Pain Was Unbearable. So Why Did Doctors Turn Her Away?

A sweeping drug addiction risk algorithm has become central to how the US handles the opioid crisis. It may only

be making the crisis worse. *Wired*. <https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/>

⁶¹ Coley, R. Y., Johnson, E., Simon, G. E., Cruz, M., & Shortreed, S. M. (2021). Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. *JAMA psychiatry*, 78(7), 726-734.

<https://pubmed.ncbi.nlm.nih.gov/33909019/>

⁶² Goggin, B. (2019, January 6). Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts. *Business Insider*.

<https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12>

⁶³ Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

<https://www.science.org/doi/full/10.1126/science.aax2342>

⁶⁴ Mateescu, A. (2021, November 16). Electronic Visit Verification: The Weight of Surveillance and the Fracturing of Care. *Data & Society*. <https://datasociety.net/library/electronic-visit-verification-the-weight-of-surveillance-and-the-fracturing-of-care/>.

⁶⁵ Allen, M. (2018, November 21). You Snooze, You Lose: Insurers Make The Old Adage Literally True.

ProPublica. <https://www.propublica.org/article/you-snooze-you-lose-insurers-make-the-old-adage-literally-true>

⁶⁶ Burke, G. and O'Brien, M. (2023, October 20). Health providers say AI chatbots could improve care. But

research says some are perpetuating racism. *AP News*. <https://apnews.com/article/ai-chatbots-racist-medicine-chatgpt-bard-6f2a330086acd0a1f8955ac995bdde4d>

- I. Loan-allocation processes,⁶⁷ financial-system access determinations,⁶⁸ credit scoring,⁶⁹ determining who is subject to a financial audit,⁷⁰ insurance processes including risk assessments,⁷¹ interest rate determinations,⁷² or financial systems that apply penalties (e.g., that can garnish wages or withhold tax returns);⁷³
- J. Decisions regarding access to, eligibility for, or revocation of government benefits or services;⁷⁴ allowing or denying access—through biometrics or other means (e.g., signature matching)—to IT

⁶⁷ Kumar, I. E., Hines, K. E., & Dickerson, J. P. (2022, July). Equalizing credit opportunity in algorithms: Aligning algorithmic fairness research with us fair lending regulation. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 357-368). <https://dl.acm.org/doi/abs/10.1145/3514094.3534154> ; Relman Colfax PLLC. (2021, April 14). Fair Lending Monitorship of Upstart Network’s Lending Model. *Initial Report of the Independent Monitor*. https://www.reلمانlaw.com/media/cases/1088_Upstart%20Initial%20Report%20-%20Final.pdf.

⁶⁸ Buolamwini, J. (2022, January 27). The IRS Should Stop Using Facial Recognition. *The Atlantic*. <https://www.theatlantic.com/ideas/archive/2022/01/irs-should-stop-using-facial-recognition/621386/>

⁶⁹ Heaven, W. D. (2021, June 17). Bias isn’t the only problem with credit scores—and no, AI can’t help. *MIT Technology Review*. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>

⁷⁰ Elzayn, H., Smith, E., Hertz, T., Ramesh, A., Goldin, J., Ho, D. E., & Fisher, R. (2023). Measuring and mitigating racial disparities in tax audits. Stanford Institute for Economic Policy Research (SIEPR). https://dho.stanford.edu/wp-content/uploads/IRS_Disparities.pdf

⁷¹ Wiggins, B. (2020). *Calculating race: Racial discrimination in risk assessment*. Oxford University Press.

⁷² Klein, A. (2019, April 11). Credit denial in the age of AI. *Brookings*. <https://www.brookings.edu/articles/credit-denial-in-the-age-of-ai/>

⁷³ Charette, R. N. (2018, January 24). Michigan 's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold. *IEEE Spectrum*. <https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold>

⁷⁴ Hao, K. (2020, December 4). The coming war on the hidden algorithms that trap people in poverty. *MIT Tech Review*. <https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/> ; Stanley, J. (2017, June 2). Pitfalls of Artificial Intelligence Decisionmaking Highlighted In Idaho ACLU Case. *ACLU*. <https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case> ; Brown, L., Richardson, M., Shetty, R., Crawford, A., and Hoagland, T. (2020, October). Challenging the use of algorithm-driven decision-making in benefits determinations affecting people with disabilities. *Center for Democracy and Technology*. <https://cdt.org/wp-content/uploads/2020/10/2020-10-21-Challenging-the-Use-of-Algorithm-driven-Decision-making-in-Benefits-Determinations-Affecting-People-with-Disabilities.pdf>

systems for accessing services for benefits;⁷⁵ detecting fraud;⁷⁶ assigning penalties in the context of government benefits;⁷⁷ or

- K. Recommendations or decisions about child welfare, child custody, or whether a parent or guardian is suitable to gain or retain custody of a child.⁷⁸
- a. Suggested changes:
 - i. in the opening phrase, change "recommendations or decisions about" to read "recommendations, decisions, or *risk assessments* about" to include child welfare risk assessments.⁷⁹
 - ii. in the phrase "IT systems for accessing services for benefits" change to "IT systems for accessing services, benefits, or *voting*."⁸⁰
 - b. Suggested additions: adoption matching,⁸¹ elder abuse and neglect,⁸² or intimate partner violence.⁸³

⁷⁵ Buolamwini, J. (2022, January 27). The IRS Should Stop Using Facial Recognition. *The Atlantic*. <https://www.theatlantic.com/ideas/archive/2022/01/irs-should-stop-using-facial-recognition/621386/>; Kenney, A. (2021, July 4). No Internet, No Unemployment: Solving This ID.me Glitch Took Two Months And A Journey Across The Rural Front Range. *CPR News*. <https://www.cpr.org/2021/07/07/colorado-unemployment-idme-glitch-internet-access/>; Kyle Wiggers. Automatic signature verification software threatens to disenfranchise U.S. voters. *VentureBeat*. Oct. 25, 2020. <https://venturebeat.com/2020/10/25/automatic-signature-verification-software-threatens-to-disenfranchise-u-s-voters/>

⁷⁶ Charette, R. N. (2018, January 24). Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold. *IEEE Spectrum*. <https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold>; Gilman, M. (2020, February 14). AI algorithms intended to root out welfare fraud often end up punishing the poor instead. *The Conversation*. <https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625>

⁷⁷ Ibid.

⁷⁸ Ho, S. and Burke, G. (2023, March 15). Here's how an AI tool may flag parents with disabilities. *AP News*. <https://apnews.com/article/child-protective-services-algorithms-artificial-intelligence-disability-02469a9ad3ed3e9a31ddae68838bc76e>; Ho, S. and Burke, G. (2023, January 31). Child welfare algorithm faces Justice Department scrutiny. *AP News*. <https://apnews.com/article/justice-scrutinizes-pittsburgh-child-welfare-ai-tool-4f61f45bfc3245fd2556e886c2da988b>; Anjana Samant, Aaron Horowitz, Kath Xu, and Sophie Beiers. Family Surveillance by Algorithm. ACLU. Accessed May 2, 2022. <https://www.aclu.org/fact-sheet/family-surveillance-algorithm>

⁷⁹ Ibid.

⁸⁰ Kyle Wiggers. Automatic signature verification software threatens to disenfranchise U.S. voters. *VentureBeat*. Oct. 25, 2020. <https://venturebeat.com/2020/10/25/automatic-signature-verification-software-threatens-to-disenfranchise-u-s-voters/>

⁸¹ Ho, S. and Burke, G. (2023, November 6). Inspired by online dating, AI tool for adoption matchmaking falls short for vulnerable foster kids. *AP News*. <https://apnews.com/article/ai-adoption-investigation-eharmony-child-welfare-f803bf3faa02bc90d285e68b1d2bc560>; Ho, S. and Burke, G. (2023, November 6). Does an AI tool help boost adoptions? Key takeaways from an AP Investigation. *AP News*. <https://apnews.com/article/ai-adoption-investigation-eharmony-child-welfare-f51f9573e3ced729a277c6817f45ffd8>

⁸² Beach, S. R., Carpenter, C. R., Rosen, T., Sharps, P., & Gelles, R. (2016). Screening and detection of elder abuse: Research opportunities and lessons learned from emergency geriatric care, intimate partner violence, and child abuse. *Journal of elder abuse & neglect*, 28(4-5), 185-216.

⁸³ Ibid.

In addition to the presumption that some use cases are rights-impacting, there are two key components of the determination of whether a system is required to follow the minimum practices that we believe deserve further clarification: the definition of AI and the waiver process.

Clarification to the definition of AI

As evidenced by agencies' drastic underreporting to OMB and the public in previous AI use case inventories,⁸⁴ which excluded many AI use cases that were publicly known, agencies need further clarification and encouragement to report all their AI use cases. As appropriately clarified in the draft memo, this includes all machine learning systems. We suggest that OMB strengthen that section of the guidance to make it clear that the inclusion of all such systems as within scope of the guidance is not simply a suggestion, but a requirement. Specifically, we suggest changing "For the purposes of this memorandum, the following technical context may assist in interpreting this definition" to read "For the purposes of this memorandum, the following technical content guides the interpretation of and extends this definition."

Waiver process

The requirements and process that agencies must satisfy to be given a waiver from instituting the minimum practices for identified safety- or rights-impacting AI systems are unclear, which may undermine the federal government's efforts to protect Americans from algorithmic harms. Currently, the draft memo states that some components of the minimum practices can be waived "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."

First, waiving *all* minimum practices in such situations is inappropriate — OMB can and should always expect agencies to test their systems for effectiveness and discrimination, among other minimum practices. We encourage OMB to add the following as minimum practices that may *not* be excluded from the requirements through the waiver process:

- Complete an AI impact assessment (Section 5.c.iv.A)
- Test the AI for performance in a real-world context (Section 5.c.iv.B)
- Conduct ongoing monitoring and establish thresholds for periodic human review (Section 5.c.iv.D)
- Mitigate emerging risks to rights and safety (Section 5.c.iv.E)
- Take steps to ensure that the AI will advance equity, dignity, and fairness (Section 5.c.v.A)
- Conduct ongoing monitoring and mitigation for AI-enabled discrimination (Section 5.c.v.C)

The above steps—appropriately identified by OMB as *minimum* practices—are necessary to ensure basic functioning, safety, and non-discrimination of any AI system. The federal government should not use AI systems that do not meet this bare minimum bar.

⁸⁴ Heilweil, R. and Adler, M. (2023, August 16) OMB acknowledges issues with process for inventorying AI use cases. *FedScoop*. <https://fedscoop.com/omb-acknowledges-issues-with-process-for-inventorying-ai-use-cases/>

Second, OMB should clarify in its final guidance that waivers require a justification and a risk assessment for *each* minimum practice for which the waiver is desired. Given that these practices are a *minimum* bar, and given that the practices already include caveats throughout such as "where appropriate," "where possible," and "to the extent practicable," the requirements for risk assessment and reporting to OMB and the public about waivers to these reasonable and necessary minimum practices should be strict. Waivers should be a tool of last resort, and should only be granted per system and per minimum practice waived, with justification for each such practice waived.

6. Do the minimum practices identified for safety-impacting and rights-impacting AI set an appropriate baseline that is applicable across all agencies and all such uses of AI? How can the minimum practices be improved, recognizing that agencies will need to apply context-specific risk mitigations in addition to what is listed?

Overall, the minimum practices identified are appropriate and necessary to ensure the safe and rights-respecting functioning of AI, although we offer some suggested edits and additions below. The requirements for impact assessments and public consultation are especially important parts of public accountability for AI, and must be coupled with robust public transparency into the methods and results of these assessments. Our specific suggestions for reporting requirements for the AI use case inventory are below in response to question 8, but we note here that impact assessments require such reporting in order to be a successful form of AI accountability.⁸⁵

Guidance throughout the draft memo identifies that if a minimum practice can not be met (for example, if a demographic disparity can not be mitigated), then agencies should not use or integrate the AI tool. These requirements that agencies not use AI tools that do not work or otherwise do not meet federal government expectations are of key importance and must not be weakened in the final guidance.

Consult and incorporate feedback from affected groups (Section 5.c.v.B.)

The draft memo's directive for federal agencies to "consult and incorporate feedback from affected groups" is a critical component of the federal government's AI governance. AI researchers increasingly have identified public participation in technology design, deployment, and oversight as a critical safeguard.⁸⁶ Research indicates that public participation, when done well, improves decision-making by incorporating the viewpoints of those most likely to be impacted by technologies.⁸⁷ Further, the memo's incorporation of public input "*before* initiating use of new or existing rights-impacting AI" is important—

⁸⁵ Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021, June 29). Assembling accountability: algorithmic impact assessment for the public interest. *Data & Society Report*.
<https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/>

⁸⁶ <https://dl.acm.org/doi/10.1145/3551624.3555290>; <https://arxiv.org/pdf/2310.00907.pdf>.

⁸⁷ Gilman, Michele E., Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society (November 2, 2022). *Fordham Law Review*, Vol. 91, 2022, University of Baltimore School of Law Legal Studies Research Paper Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4266250>; Michele Gilman, *Democratizing AI: Principles for Meaningful Public Participation*, Data & Society (Sept. 27, 2023), <https://datasociety.net/library/democratizing-ai-principles-for-meaningful-public-participation/>.

for participation to be meaningful, affected groups should be involved at the beginning of the process, particularly in establishing the desirability or need of the system in the first place.

However, as currently drafted, the guidance only requires federal agencies to consult affected groups “to the extent practicable and consistent with applicable law and governmentwide guidance,” leaving a substantial loophole for federal agencies to bypass public input. Specifically, many federal agencies—already facing capacity and staffing constraints—may find that additional procedures such as holding public meetings, gathering public input, soliciting comments through an RFI, and consulting with federal employees’ unions on the use of AI to fall beyond “the extent practicable.”

The key issue with public participation is that *it works only when done well*.⁸⁸ Under the memo as drafted, many federal agencies may either ignore the directive, or go through rote motions of cursory listening sessions and RFIs without meaningfully elevating the voice of marginalized communities. The guidance could be strengthened by omitting “To the extent practicable,” which would encourage robust adoption of participatory methods while still preserving the condition that agencies’ public consultation be “consistent with applicable law and government guidance.”

Missing practices

Two key principles of the Biden-Harris administration's Blueprint for an AI Bill of Rights⁸⁹ are missing or only partly satisfied by the minimum practices of the draft memorandum: explanation and data privacy.

Data privacy

Excluding data privacy from the minimum practices entirely misses a major opportunity to safeguard people's rights because, as the draft memo identifies, data is a key part of any AI system. Given OMB's role in overseeing compliance with the Paperwork Reduction Act, this also misses an opportunity for further oversight and/or consolidation. The key minimum practice to add related to data privacy is data minimization, which is also a key pillar of the bipartisan American Data Privacy and Protection Act.⁹⁰ Specifically, we suggest adding the following minimum practice from the Biden-Harris administration's Blueprint for an AI Bill of Rights to the final guidance:

***Data collection and use-case scope limits.** Data collection should be limited in scope, with specific, narrow identified goals, to avoid “mission creep.” Anticipated data collection should be determined to be strictly necessary to the identified goals and should be minimized as much as possible. Data collected based on these identified goals and for a specific context should not be used in a different context without assessing for new privacy risks and implementing appropriate mitigation measures, which may include express consent. Clear timelines for data retention should be established, with data deleted as soon as possible in accordance with legal or policy-based limitations. Determined data retention timelines should be documented and justified.*

⁸⁸ Ibid.

⁸⁹ The White House. (2022, October). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

⁹⁰ 117th Congress. (2022). H.R. 8152 American Data Privacy and Protection Act. <https://www.congress.gov/117/bills/hr8152/BILLS-117hr8152rh.pdf>

Explanation

Explanations, especially of adverse decisions about individuals, are necessary in order for those individuals to seek meaningful recourse. Explanations are broadly understood as important and form a core pillar of Leader Schumer's SAFE Innovation Framework for future AI legislation.⁹¹ Additionally, such explanations are already required under federal law in the case of adverse action notices by financial institutions—even in cases where AI informs decisions—as made clear by the CFPB.⁹² Explanations as required by the CFPB must be specific and "accurately indicate the principal reason(s) for the adverse action."⁹³ Such explanations are useful to people and, importantly, *are technically feasible to generate* no matter the type of AI system. Footnote 37 of the draft OMB memorandum states that "exact explanations of AI decisions are often not technically feasible." While it's true that an "exact" explanation would require a re-explanation of the entire AI system and is thus not *useful* to a person, it's certainly possible to generate. More importantly, meeting the standard set by the CFPB for AI system explanations is both useful and technically feasible. Common methods⁹⁴ have readily available software packages⁹⁵ that can be used to generate feature importance values that are model-agnostic. We suggest adding the following minimum practice to the continuing practices part of Section 5.c.v, adapted from OMB's draft memorandum footnote 37:

***Explanations.** Agencies must provide explanations to individuals when AI meaningfully influences adverse actions by rights-impacting AI systems which specifically impact them. Explanations provided to the individual shall: explain why the AI system produced the result it produced for this specific individual; be scientifically valid, meaningful, useful, and as simply stated as possible; accurately indicate the principal reason(s) for the adverse action; and include more comprehensive explanations for higher-risk decisions.*

7. What types of materials or resources would be most valuable to help agencies, as appropriate, incorporate the requirements and recommendations of this memorandum into relevant contracts?

⁹¹ Schumer's SAFE Innovation Framework https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf

⁹² The Consumer Financial Protection Bureau. (2023, September 19). Adverse action notification requirements and the proper use of the CFPB's sample forms provided in Regulation B. *Consumer Financial Protection Circular 2023-03*. <https://www.consumerfinance.gov/compliance/circulars/circular-2023-03-adverse-action-notification-requirements-and-the-proper-use-of-the-cfpbs-sample-forms-provided-in-regulation-b/>

⁹³ The Consumer Financial Protection Bureau. (2023, September 19). Adverse action notification requirements and the proper use of the CFPB's sample forms provided in Regulation B. *Consumer Financial Protection Circular 2023-03*. <https://www.consumerfinance.gov/compliance/circulars/circular-2023-03-adverse-action-notification-requirements-and-the-proper-use-of-the-cfpbs-sample-forms-provided-in-regulation-b/>

⁹⁴ Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. <https://dl.acm.org/doi/10.5555/3295222.3295230> ; Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). <https://dl.acm.org/doi/abs/10.1145/2939672.2939778>

⁹⁵ SHAP python package <https://github.com/shap/shap>; LIME python package <https://github.com/marcotcr/lime>

Concern that the values language in the draft memorandum does not center equity

We are concerned that the procurement section (Section 5.d) of the draft memorandum introduces new principles focused on values in the particular context of procurement that do not align with the requirements in Executive Order 14110. The values language included in this section does not address the Administration's existing equity commitments. Specifically, the procurement section copies language from Executive Order 13960, which says that AI should be:

Lawful and respectful of our Nation's values. Agencies shall design, develop, acquire, and use AI in a manner that exhibits due respect for our Nation's values and is consistent with the Constitution and all other applicable laws and policies, including those addressing privacy, civil rights, and civil liberties.

The procurement section of the draft memorandum similarly says:

Aligning to National Values and Law. Agencies should ensure that procured AI exhibits due respect for our Nation's values, is consistent with the Constitution, and complies with all other applicable laws, regulations, and policies, including those addressing privacy, confidentiality, copyright, human and civil rights, and civil liberties.

Yet Executive Order 14110 clearly states that:

Artificial Intelligence policies must be consistent with my Administration's dedication to advancing equity and civil rights. My Administration cannot—and will not—tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice. From hiring to housing to healthcare, we have seen what happens when AI use deepens discrimination and bias, rather than improving quality of life. Artificial Intelligence systems deployed irresponsibly have reproduced and intensified existing inequities, caused new types of harmful discrimination, and exacerbated online and physical harms. My Administration will build on the important steps that have already been taken—such as issuing the Blueprint for an AI Bill of Rights, the AI Risk Management Framework, and Executive Order 14091 of February 16, 2023 (Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government)—in seeking to ensure that AI complies with all Federal laws and to promote robust technical evaluations, careful oversight, engagement with affected communities, and rigorous regulation. It is necessary to hold those developing and deploying AI accountable to standards that protect against unlawful discrimination and abuse, including in the justice system and the Federal Government. Only then can Americans trust AI to advance civil rights, civil liberties, equity, and justice for all.

and Executive Order 14091 states that:

When designing, developing, acquiring, and using artificial intelligence and automated systems in the Federal Government, agencies shall do so, consistent with applicable law, in a manner that advances equity.

The Biden-Harris administration's commitment to equity, as evidenced through many executive orders (including 13985, 14091, 14110, and others) and administration practices, go beyond a minimal commitment to upholding existing civil rights law as stated in the draft memorandum. The minimum

practices identified by OMB in the draft also ensure more than this minimum standard. **We strongly encourage OMB to revise these principles in the context of procurement and follow the clear directives in favor of equity from this administration; as written, the procurement principles do not meet the expectations of EOs 14091 or 14110.**

Contract requirements that include the minimum practices

There should be no difference between the requirements for agency use of AI based on whether the agency procured the system or developed it in-house; contract requirements should be based on the minimum practices detailed in the OMB draft memorandum. The process of determining whether a system is safety- or rights-impacting is already detailed in the draft memorandum and should also be followed for procured systems. Similarly, the AI use case inventory should include procured systems.

Contracts can incorporate these requirements by:

1. **Instituting the minimum practices** of the draft memorandum as required clauses in any safety-impacting or rights-impacting AI system contract;
2. Simultaneously **procuring an external evaluator** to conduct independent testing and monitoring of the AI system, for example using the existing JAIC contract vehicle⁹⁶; and
3. **Requiring detailed reporting** via the AI use case inventory as part of both contracts.

We encourage the Office of Federal Procurement Policy (OFPP) to build on the strong foundation laid by the minimum practices described in the draft memorandum—these practices can and should be applied to procurement as well. There is no need to reinvent the wheel. The procurement section of the draft memorandum introduces new practices focused on generative AI, but the draft memorandum already covers generative AI within its definition of AI. Procurement guidelines should follow this approach.

Detailed reporting and the AI use case inventory

Including the detailed reporting via the AI use case inventory is an important part of public accountability for these systems and will also serve to provide contracting agents with information they need to appropriately assess a system to determine whether it is safe, effective, and rights-protecting. **We thus encourage OFPP to examine the responses to question 8 of this RFI;** there may be cases where OMB decides not to incorporate detailed questions and responses into the public-facing portion of the AI use case inventory, yet these questions can provide a useful starting point for contracting officers aiming to privately assess a system. A similar approach that relies on an assessment of detailed questions answered by a vendor is used by an industry consortium to assess HR vendor systems that can include AI.⁹⁷

Federal grants must also require the minimum practices

Finally, we encourage OMB's Office of Federal Financial Management to work with OFPP to learn from this procurement process and incorporate similar protections into their grant guidance. Federal grants

⁹⁶ JAIC Public Affairs. (2022, February 11). AIC Offers New Enterprise-Wide Contract Vehicle for Rapid Procurement of AI Test & Evaluation. https://www.ai.mil/blog_02_11_22_jaic_new_contract_vehicle.html

⁹⁷ Data & Trust Alliance. Algorithmic Safety: Mitigating Bias in Workforce Decisions. <https://dataandtrustalliance.org/our-initiatives/algorithmic-safety-mitigating-bias-in-workforce-decisions>

have a substantial impact on state and local governments and communities. The federal government should not be providing monetary support to instituting AI systems that would violate the government's own guidelines and present risks to the public's safety or rights.

8. What kind of information should be made public about agencies' use of AI in their annual use case inventory?

Public reporting on agencies' AI uses—including impact assessments, risk mitigation procedures, and demographic disparity measurements and results, among other items—are key to effective AI accountability.⁹⁸ OMB's plans to integrate the collection of this data into its annual Integrated Data Collection process represent an important step, and ensuring that information about *all* safety-impacting and rights-impacting AI systems is reported publicly will be critical for building public trust in these systems.

Including all AI systems in the inventory

As also described above in response to question 5, previous AI use case inventories have not included all AI systems;⁹⁹ as part of this renewed AI use case inventory, OMB must close existing loopholes and make clear to agencies that *all* AI use cases must be reported. For example, in the existing AI use case inventory, the Department of Justice reported only 4 AI use cases,¹⁰⁰ but did not include the facial recognition services that GAO has reported¹⁰¹ are used by DOJ, nor did they include the DOJ-developed PATTERN risk assessment tool.¹⁰² Two possible loopholes that should be closed by OMB include:

1. Lack of reporting of commercially developed software used by the federal government through acquisition, and
2. Lack of reporting of tools perceived to be less sophisticated than cutting-edge AI systems.

Given that the OMB guidance applies to agency "use of AI" and is clearly meant to apply to acquired systems as well, OMB should make clear to agencies that they are required to report on their use of commercially developed AI, such as the facial recognition tools identified by GAO as in use by DOJ. Additionally, some agencies may be under-reporting AI systems because they interpret trained regression

⁹⁸ Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021, June 29). Assembling accountability: algorithmic impact assessment for the public interest. *Data & Society Report*.

⁹⁹ Heilweil, R. and Adler, M. (2023, August 16) OMB acknowledges issues with process for inventorying AI use cases. *FedScoop*. <https://fedscoop.com/omb-acknowledges-issues-with-process-for-inventorying-ai-use-cases/>

¹⁰⁰ Department of Justice. AI Use Case Inventory Submission on Open Data. <https://www.justice.gov/open/page/file/1517316/download>

¹⁰¹ U.S. Government Accountability Office. (2023, September 12). Facial Recognition Services: Federal Law Enforcement Agencies Should Take Actions to Implement Training, and Policies for Civil Liberties. <https://www.gao.gov/products/gao-23-105607>

¹⁰² U.S. Department of Justice. (2021, December). 2021 Review and Revalidation of the First Step Act Risk Assessment Tool. *NIJ report NCJ 303859*. <https://www.ojp.gov/pdffiles1/nij/303859.pdf>; Ryan Labrecque. (2023, March). 2022 Review and Revalidation of the First Step Act Risk Assessment Tool. *NIJ report NCJ Number 305720*. <https://nij.ojp.gov/library/publications/2022-review-and-revalidation-first-step-act-risk-assessment-tool>

tools (like those that make up PATTERN¹⁰³) as not being within scope of the definition of artificial intelligence under the 2019 NDAA, as also used in the OMB draft memorandum. But linear regression, logistic regression, and other such models *are* machine learning models,¹⁰⁴ which should be considered within the scope of the definition of artificial intelligence used by OMB since (as OMB rightly notes in its definitional clarification in the draft memorandum) the definition encompasses machine learning. We suggest that OMB make this needed clarification by changing "For the purposes of this memorandum, the following technical context may assist in interpreting this definition" to read "For the purposes of this memorandum, the following technical content guides the interpretation of and extends this definition"—and make it clear that agencies are expected to fully report their AI use cases.

Summary reporting of excluded AI systems

Given the waiver process and that some systems are additionally excluded from public reporting as part of the AI use case inventory, it is important that there be public reporting summarizing these exclusions. Such reporting should include at least the following information:

- Number of AI systems excluded from public reporting per agency and per rationale (whether excluded via Section 5.c.i or via a different mechanism);
- Number of AI systems subject to an extension for compliance with minimum practices per agency (Section 5.c.ii);
- Detailed justifications and plans to meet the minimum requirements per AI system that has been granted an extension for compliance with minimum practices (Section 5.c.ii);
- Number of waivers granted per agency and per minimum practice (Section 5.c.iii); and
- The risk assessment done to justify each granted waiver or detailed justification of agency need for critical operations (Section 5.c.iii).

Detailed reporting of testing and risk as part of the AI use case inventory

For each minimum practice described in the draft memo, the public should receive information and assurances through the AI use case inventory that the agency took the required protective steps. This should include providing specifics about risk mitigation practices and resulting findings. Example questions for a few key minimum practices are included below, but a more comprehensive exercise should be undertaken to ensure detailed and specific reporting for each practice. The below suggested questions for the AI use case inventory draw directly from the documentation suggested by and needed to meet OMB's draft memo list of minimum practices.

Complete an AI impact assessment (Section 5.c.iv.A)

The intended purpose for the AI and its expected benefit (Section 5.c.iv.A.1)

- What is the intended purpose of the AI system?
- What is the expected benefit of the AI system?

¹⁰³ U.S. Department of Justice. (2021, December). 2021 Review and Revalidation of the First Step Act Risk Assessment Tool. *NIJ report NCJ 303859*. <https://www.ojp.gov/pdffiles1/nij/303859.pdf>; Ryan Labrecque. (2023, March). 2022 Review and Revalidation of the First Step Act Risk Assessment Tool. *NIJ report NCJ Number 305720*. <https://nij.ojp.gov/library/publications/2022-review-and-revalidation-first-step-act-risk-assessment-tool>.

¹⁰⁴ Russell, S. and Norvig, P. (2022) Artificial Intelligence: a modern approach. 4th U.S. edition. <https://aima.cs.berkeley.edu/>, see Chapter 19.6 on linear regression <https://aima.cs.berkeley.edu/contents.html>

- What performance metrics and/or qualitative analysis was used to assess the fitness to purpose and expected benefit of the AI system?
- What were the quantitative and/or qualitative results found by the above described analysis?

The potential risks of using AI (Section 5.c.iv.A.2)

- What stakeholders will be most impacted by the use of the AI system?
- What possible failure modes may result from use of the AI and of the broader system, both in isolation and as a result of human users and other likely variables outside the scope of the system itself?
- What are the potential risks to underserved communities?
- Describe how the expected benefits of the AI functionality were weighed against its potential risks and how it was determined that the benefits meaningfully outweigh the risks. Note that if the benefits do not meaningfully outweigh the risks, agencies should not use the AI.

The quality and appropriateness of the relevant data (Section 5.c.iv.A.3)

- What is the provenance and quality of the data for its intended purpose?
- How is the data relevant to the task being automated and why does it have a reasonable expectation of being useful for the AI's development, testing, and operation?
- Does the data contain sufficient breadth to address the range of real-world inputs the AI might encounter? How was that assessed?
- Does the data come from an adequately reliable source? How was that assessed?
- How are errors from data entry, machine processing, or other sources measured and limited? What are the associated data error rates? This should include errors from relying on AI-generated data as training data or model inputs.

Take steps to ensure that the AI will advance equity, dignity, and fairness (Section 5.c.v.A)**Proactively identifying and removing factors contributing to algorithmic discrimination or bias (Section 5.c.v.A.1)**

- What is the process by which factors were assessed for bias?
- List any factors which were proactively removed based on this assessment.

Assessing and mitigating disparate impacts (Section 5.c.v.A.2)

- What metrics and/or qualitative analysis was used to assess disparities in the AI's performance across demographic groups?
- What demographic groups were included in the assessment?
- How was demographic data acquired or inferred to perform the assessment?
- What were the quantitative and/or qualitative results found by the above described analysis per demographic group?
- How was the AI system's real-world deployment impact assessed, and what did that assessment find?
- How were identified disparities that have the potential to lead to discrimination, cause meaningful harm, or decrease equity, dignity, or fairness mitigated? Note that if adequate mitigation of the disparity is not possible, then agencies should not use or integrate the AI tool.

Using representative data (Section 5.c.v.A.3)

- What communities will be affected by the AI system?
- What demographic groups are included in the data used to develop, operate, and assess the AI system?
- How was demographic data acquired or inferred to perform the assessment?
- What is the total number of people represented in the training data? What is the total number of people represented in the test data?
- What is the number of people from each of the demographic groups in the training data and in the test data?
- How was the data used to develop, operate, and assess the AI collected?
- What is the historical and societal context of the data and data collection process?
- Describe how any improper bias in the data has been assessed and mitigated.

Consult and incorporate feedback from affected groups (Section 5.c.v.B)

- On what date(s) was consultation with affected groups, including underserved communities, conducted about the design, development, and use of the AI?
- What affected groups were consulted? Affected groups may include customers, federal employee groups, and employees' union representatives among other members of the public.
- What type of consultation was conducted? Choose from the following (you may choose more than one):
 1. Direct user testing, such as observing users interacting with the system;
 2. General solicitations of comments from the public, such as a request for information in the Federal Register or a "Tell Us About Your Experience" sheet with open-ended space for responses;
 3. Post-transaction customer feedback collections;
 4. Public hearings or meetings, such as a listening session; or
 5. Any other transparent process that seeks public input, comments, or feedback from the affected groups in a meaningful, equitable, accessible, and effective manner.
- If the consultation listed above was some other transparent process (option 5), describe that process.
- If consultation with affected groups was not conducted because doing so was inconsistent with applicable law or governmentwide guidance, describe the applicable law or governmentwide guidance that is inconsistent with such consultation.
- Provide a URL to any public documentation of the consultation, such as resulting RFI submissions.
- Give a summary of the key findings of the consultation.
- Describe any changes made to the AI system and/or planned rollout in response to the consultation.

Respectfully submitted,

Sorelle Friedler, Senior Policy Fellow

Brian J. Chen, Policy Director