CYBERLAW CLINIC
HARVARD LAW SCHOOL | BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY

To: The Honorable Dominic J. Mancini
Acting Administrator,
Office of Information and Regulatory Affairs
Office of Management and Budget

From: Commenters (Defined Below)

Date: March 13, 2020

Re: Draft Memorandum to the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Applications

---

INTRODUCTION

The Cyberlaw Clinic at Harvard Law School submits this comment on behalf of:

Amar Ashar
Assistant Research Director,
Berkman Klein Center for Internet & Society;

Christopher Bavitz
WilmerHale Clinical Professor of Law, Harvard Law School
Managing Director, Cyberlaw Clinic
Faculty Co-Director, Berkman Klein Center for Internet & Society;

Ryan Budish
Assistant Research Director,
Berkman Klein Center for Internet & Society;

Jessica Fjeld
Lecturer on Law at Harvard Law School
Assistant Director of the Cyberlaw Clinic;

Mason Kortz
Clinical Instructor, Cyberlaw Clinic; and

Adam Nagy
Project Coordinator, Berkman Klein Center for Internet & Society;

(collectively, "Commenters").[1] Commenters have substantial expertise with respect to regulatory, ethics, and rights-based approaches to Artificial Intelligence ("AI").

Notably, several of the Commenters recently contributed to a mapping of growing global consensus on the regulation of development and deployment of AI systems in both the public and private sectors.[2] Commenters have engaged with governments on topics relating to AI and produced significant academic scholarship with respect thereto.[3]

OVERVIEW

Commenters appreciate the opportunity to offer their perspective on the Office of Management and Budget (the "OMB")'s "Draft Memorandum to the Heads of Executive Departments and Agencies, Guidance for Regulation of Artificial Intelligence Applications" (the "Draft Memorandum"). Specifically, Commenters wish to commend the Draft Memorandum's focus on narrow ("weak") AI and its promotion of performance-based analyses and stakeholder engagement while also providing their expertise on how the Draft Memorandum's guidance can better balance the competing

---

[1] Commenters write in their individual capacities; institutional affiliations are provided for identification purposes only.

[2] Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, & Madhulika Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, BERKMAN KLEIN CENTER RESEARCH PUBLICATION NO. 2020-1 (January 15, 2020), https://ssrn.com/abstract=3518482 & http://dx.doi.org/10.2139/ssrn.3518482.

[3] Commenters thank spring 2020 Harvard Law School Cyberlaw Clinic students João Marinotti (HLS JD '20) and Jonathan Iwry (HLS JD '20) for their significant contributions to this comment.

interests and values promoted in Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence" (The "Executive Order").[4]

As the Executive Order notes, the United States "must drive technological breakthroughs" and "reduce barriers" to the development and deployment of safe and just AI technologies. At the same time, the United States must "foster public trust and confidence in AI" and must "protect civil liberties, privacy, and American values," including the principles of freedom, human rights, the rule of law, and respect for intellectual property. This comment aims to demonstrate how the OMB's Draft Memorandum could better address the potential conflict between these goals. Although some degree of tension (and a resulting need for tradeoffs) ultimately might be unavoidable, we think it possible to strike an optimal balance among these overlapping values in serving the United States' economic and civil-democratic interests.

The Comment is divided into three Sections. Section 1 provides the necessary economic and social context to demonstrate that regulation and innovation are not mutually exclusive; rather, some forms of Federal regulation on AI have explicitly been requested by the private sector to facilitate innovation. Section 2 evaluates the Draft Memorandum's promotion of cost-benefit (or benefit-cost) analysis ("CBA") to demonstrate that more guidance is necessary if administrative agencies are meant to effectively employ CBA in their regulatory and non-regulatory actions surrounding AI. Section 3 turns to the 10 Principles itemized in the Draft Memorandum and provides concrete commentary on how the information in Sections 1 and 2 should affect the language and content of the Draft Memorandum.

SECTION 1 - CONTEXT OF AI REGULATION

THE PRIVATE SECTOR WANTS AI-SPECIFIC REGULATION

The Draft Memorandum aims to ensure that "American companies are not disadvantaged by the United States' regulatory regime." It hopes to prevent Federal actions that "needlessly hamper AI innovation and growth."

[4] United States, Executive Office of the President Donald J. Trump. Executive Order 13859: Maintaining American Leadership in Artificial Intelligence.

By presenting regulation and growth as mutually exclusive, the Draft Memorandum disregards the growing private sector consensus that Federal regulation is, in fact, necessary for continued American leadership in artificial intelligence. American companies such as Microsoft[5], Google (Alphabet)[6], and IBM[7], as three examples, have noted their desire for Federal regulation of AI technologies. Without Federal AI regulation, American companies may face competing and contradictory state regulatory regimes and may encounter a patchwork of more stringent foreign laws, created in the absence of American regulation, as occurred in the privacy realm under the California Consumer Privacy Act[8] and the European Union's General Data Protection Regulation[9], respectively.

Informed Federal AI regulation will aid American technology companies in avoiding legal and public-relations crises. Without clear Federal guidance, developers, deployers, and investors of AI technologies must make decisions under the shadow of liability in an uncertain legal landscape. Lack of regulation and education may lead to crises for American companies. Facebook Inc. ("Facebook") has gained notoriety for its data scandal involving Cambridge Analytica, in which the personal information of millions of Facebook users was shared with an external political consulting firm without those users' consent. By educating and collaborating with stakeholders in the AI industry during the regulatory process, developers, deployers, and investors of AI can work with legal certainty in the face of growing public scrutiny of technology companies.

For example, micro-targeted advertising, which uses weak AI as defined in the Draft Memorandum, has led companies such as Facebook to

---

[5] Ryan Hagemann & Jean-Marc Leclerc (Co-Directors of IBM Policy Lab), *Precision Regulation for Artificial Intelligence*, IMB POLICY LAB (Jan. 21, 2020), https://www.insurancejournal.com/news/national/2020/01/22/555445.htm.
[6] Sunday Pichai (CEO of Alphabet Inc. and Google LLC.), *Why Google thinks we need to regulate AI*, FINANCIAL TIMES (Jan. 20, 2020), https://www.ft.com/content/34676594-386d-11ea-ac3c-f68c10993b04.
[7] Brad Smith (President and Chief Legal Officer of Microsoft Corp.), *Facial recognition: It's time for action,* MICROSOFT ON THE ISSUES (Dec. 6, 2018), https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/.
[8] Cal. Civ. Code §§ 1798.100-1798.199 (2018).
[9] Regulation 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. L 119.

encounter civil rights lawsuits and administrative proceedings stemming from age, sex, and racial discrimination. Proper regulatory guidance could have prevented Facebook from allowing its AI to be used in discriminatory ways. Similarly, proper Federal regulatory guidance on AI can help other technology companies to avoid civil suits, administrative proceedings, and public backlash.

EFFECTIVE AI REGULATORY REGIMES FOR SYSTEMIC RESILIENCE

As noted, the Draft Memorandum aims to ensure that "American companies are not disadvantaged by the United States' regulatory regime," but it focuses solely on ensuring a positive regulatory environment for AI developers and deployers. It is noteworthy, though, that unregulated AI innovation could come at the cost of not only public values but also private sector economic security.

Informed Federal regulation requires a systems approach to analyzing the goals of the multi-stakeholder ecosystem that is the private sector. Regulation of AI can be used to promote economic growth by addressing market failures and ensuring a level playing field among competing private actors. Intellectual property protection and consumer confidence, for example, must not be sacrificed for the benefit of AI innovation, as such tradeoffs would do more harm than good.

Therefore, agencies should not only focus on "preventing bad actors from exploiting AI system weaknesses … and adversarial use of AI," which addresses the immediate developers and deployers of AI, but also understand that "systemic resilience" encompasses the entire private sector ecosystem surrounding AI. A lack of sufficient intellectual property, privacy, and civil rights protections will decrease the overall systemic resilience of the American technology sector and will ultimately harm American leadership in AI itself.

To that end, agencies should not view "protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property" as "barriers" to AI

innovation; rather, these form the foundation on which American AI leadership has been created and will be maintained.

SECTION 2 - APPLYING BENEFIT-COST ANALYSIS TO AI REGULATION

Appendix A to the Draft Memorandum (the "Appendix" or "Appendix A") notes that agencies "should conduct a benefit-cost analysis that estimates the benefits and costs associated with each alternative" to regulatory or non-regulatory approach to AI. To the extent possible, the "benefits and costs should be quantified and monetized," though qualitative descriptions are allowed when "quantification of a particular benefit or cost is not possible."

If CBA is the analytical tool prescribed, it is imperative that the tool be used in such a way as to achieve the systemic resilience and necessary regulatory environment envisioned in Section 1. As written, the Draft Memorandum does not adequately address the fact that, especially for AI applications, many of the factors involved are not quantifiable with a sufficient degree of certainty. Furthermore, many of the variables necessary for employing CBA will ultimately rely on a highly subjective and normative balance between competing values. Our ability to determine what counts as a benefit or cost depends on having a clear normative framework of underlying, well-defined values, interests, and goals. And even with clear objectives in place, empirical analysis depends on the ways in which empirical evidence is interpreted, and therefore can still involve a considerable degree of subjectivity and normativity. CBA is a powerful analytical tool when it is applied carefully. The way we conceptualize our goals, as well as the way we interpret evidence in light of those goals, matter greatly for our ability to conduct and apply CBA effectively.

NORMATIVITY & SUBJECTIVITY ARE NOT AVOIDED THROUGH CBA

The "regulatory impact" and the "benefit-cost" analyses described in Appendix A may lead agencies to attempt to perform a strictly empirical inquiry. In the context of AI, however, capturing the collateral benefits and costs, or internalizing all externalities, of AI applications, may prove extremely difficult. Furthermore, not all values can be converted into comparable empirical metrics. How, for example, can we put a price tag on

access to privacy? Can we adequately quantify the unknown future harms caused by privacy infringement?

The Draft Memorandum acknowledges this problem and allows agencies to describe risks qualitatively, evaluating "impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, and personal freedom." What the Draft Memorandum does not address, however, is how agencies should incorporate these evaluations into a decision-making framework alongside the promoted quantitative CBA.

This lack of guidance is especially troubling given that values such as privacy, equity, and balanced protections for intellectual property may ultimately come into conflict with each other. Furthermore, a surface-level description of the costs and benefits to each value might not reflect the total social, cultural, and economic emphasis that the government or the public places on each value.

Accordingly, an effective CBA requires that regulatory agencies determine how best to weigh the priority of their varying objectives so as to determine when a cost to one interest is offset by a benefit to a competing interest. The lack of guidance on how to properly accomplish this task, leaves room for special interests, cognitive biases, and other factors that can ultimately lead to erroneous determinations of benefits and costs.

Bright line rules relating to American values, including civil liberties, privacy, and balanced protections for intellectual property, would help to alleviate these issues and may facilitate agency decision making in future regulatory and non-regulatory approaches to AI.

THE COSTS OF GETTING CBA WRONG

As noted, the potential risks of implementing AI applications are not easy to discern or anticipate. Without taking necessary precautions, erroneous applications of CBA will inevitably lead to the adoption of harmful AI tools. Risks will be borne not only by those whose rights are affected, and costs will be imposed on those responsible for violating those rights. It is not obvious that less regulation of AI specifically would decrease the risk of legal liability. On the contrary, existing causes of action within

already-regulated industries will (appropriately) give those impacted a means of redress. And new regulation may be required to ensure those developing AI applications understand the rules of the road. Indeed, a clear and well-defined regulatory framework might help private actors to better identify the legal contours of using AI in the marketplace, reduce overall ambiguity, and navigate the economic landscape more effectively.

Implementations of CBA without adequate precautions have already had disastrous consequences both domestically and abroad. Domestically, lack of AI regulation may have contributed to Facebook's inattentive implementation of AI systems that engaged in age, sex, and racial discrimination in the context of micro-targeted housing and employent advertisement.[10] Abroad, Facebook's AI-controlled systems of content promotion and content moderation failed to prevent the proliferation of violence-inducing posts, leading to what United Nations officials called "a textbook example of ethnic cleansing."[11]

Given the risk of such adverse consequences, it is worth considering whether certain rights-based approaches in the manner of bright lines might ultimately provide greater benefit for the AI industry than unclear and uncertain applications of CBA. Bright line approaches might also protect against the sorts of cognitive biases typical of CBA. When engaging in value tradeoffs, interested parties can easily underestimate costs and overestimate benefits, especially in the context of innovation. Absolute thresholds could provide a sturdy, unambiguous reference point to counteract that tendency.

This would not be without precedent; United States tort law often invokes bright line rules when effective CBA is impracticable or insufficient to protect against serious dangers. Modern products liability cases, for instance, often apply "strict liability" against defendants responsible for defective products. The policy rationales behind such a bright line approach

---

[10] For a legal analysis, see Joseph Blass, *Algorithmic Advertising Discrimination*, 114 NW. U. L. REV. 415 (2019). For a technical analysis, see Muhammad Ali, Piotr Sapiezynski, et al., *Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes*, 3(CSCW) PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1 (2019).

[11] *UN human rights chief points to 'textbook example of ethnic cleansing' in Myanmar*, UN NEWS (Sept. 11, 2017), https://news.un.org/en/story/2017/09/564622-un-human-rights-chief-points-textbook-example-ethnic-cleansing-myanmar.

include encouraging safety in research and development,[12] which may apply equally in the realm of research and development of AI. Such uses of bright line rules elsewhere in American law suggest that lawmakers and policymakers do and should continue to value measures beyond simple CBA, especially when effective analyses are not practicable, and could offer a touch point for careful efforts to regulate AI.

During regulatory impact analyses, agencies should refer to previously determined thresholds and floors for critical American values. If AI applications would require the sacrifice of such values or violate such floors, no amount of economic benefit would outweigh these considerations. Clear constraints such as these would serve as checks on the alignment of subjective variables used in CBA with the values of the American government and the American public. Such guideposts are not hindrances to innovation; rather, they merely convey that considerations pertaining to justice might be more fragile or volatile than marginal increases in economic efficiency.

SECTION 3 - APPLICATION TO DRAFT MEMORANDUM

Commenters now turn to the 10 Principles itemized in the Draft Memorandum. In light of the discussion above, the following analyses provide guidance on how the OMB's Memorandum could better fulfill its stated goals.

PRINCIPLE 1 - PUBLIC TRUST IN AI

Many public and private sector institutions place heavy emphasis on the importance of trust. Implementing trustworthy technologies depends partly on the process of building technical systems, but it also depends on documenting, evaluating, and explaining to and for the public how important principles are put into practice. The manner in which trust is built into AI-based systems is not just a matter of technology; it also depends on economic, social, and political considerations.

---

[12] David W. Leebron, *An Introduction to Products Liability: Origins, Issues and Trends*, 1990 ANN. SURV. AM. L. 395, 397 (1991).

Good practices involving transparency and disclosure are indispensable to the process of building public trust. But while disclosure of information can often illuminate, it can also add to existing confusion and misinformation in the absence of proper context. This is especially true with respect to new and specialized technologies on the frontiers of computer science. More information might not be productive without explanation; transparency depends on intelligibility.

Government actors should generally embrace transparency while making sure to take context into account. They would also benefit from taking into account other methods for promoting public trust. These might include more thoroughly evaluating procurement procedures for AI technologies (especially in the public sector), conducting audits of algorithms, and finding other ways to make AI outputs intelligible to the general public. For example, those interested in promoting greater transparency might look for ways to clarify the nature and inner workings of AI systems—including features such as intended purpose, goals, data sources, limitations, variables used for training, and common standards by which to track development and success.

Meaningfully promoting public trust in AI requires coupling transparency with other activities. Empirical research, public impact assessments, civil sector impact reports, and multi-stakeholder mechanisms are just as important as disclosure by private entities.[13]

PRINCIPLE 2 - PUBLIC PARTICIPATION

We commend the OMB for acknowledging the value of public participation as a means of fostering trust and legitimacy. There certainly is reason to believe "public participation . . . will improve agency accountability and regulatory outcomes, as well as increase public trust and confidence." It is worth considering how best to foster public participation, and whether reliance on traditional rulemaking procedures should be

[13] For notable efforts at promoting transparency toward enhanced public trust in technologies, see MIT's Data Nutrition Project (https://datanutrition.org/) and Google's "Model Cards" initiative (https://modelcards.withgoogle.com/about).

accompanied by other ways for members of the public to understand and hold agencies accountable for uses of AI.

A key problem is that the private parties most likely to participate in the rulemaking process might be large, powerful entities, including many with significant commercial interests. A process like this one is less likely to account for the perspectives of, say, people who might be subjected to risk assessment programs or biased algorithms related to housing. Appendix A to the Draft Memorandum states that "[t]he informal rulemaking process under the Administrative Procedure Act provides predictable and meaningful opportunities for interested stakeholders to provide input on draft regulations and scrutinize the evidence and analytic bases of regulatory proposals." But not all stakeholders are equally well-equipped to navigate the legal landscape. Put simply, relying on the informal rulemaking process alone runs a risk of disproportionately selecting for and representing the interests of powerful entities that have the resources and technical know-how to access that process.

The Appendix further states that "[i]n soliciting public input on Notices of Proposed Rulemaking (NPRMs) that relate to AI applications, agencies will benefit from the perspectives and expertise of stakeholders engaged in the design, development, deployment, operation, and impact of AI applications, and facilitate a decision making process that is more transparent and accountable." The question is how best to include and incorporate the perspectives of stakeholders whose interests might be affected, but who face practical or personal roadblocks to participating in the rulemaking process. Without ensuring that parties with less access to or familiarity with the legal system are well represented, efforts to increase participation by relying heavily on the rulemaking process will skew the results in ways that disproportionately burden parties with less economic, social, or political influence.

This ties into questions about how private parties can feel confident that they will be able to hold agencies accountable, particularly in cases where their civil liberties might be jeopardized by an agency's particular decisions as to how it uses AI-related technology. One useful option to counteract the risk-favoring style of tradeoffs, to the extent that it might threaten civil liberties, would be to ensure that agencies make procedures

available that give adversely affected private parties ways of demanding compensation. Regarding this approach, if affected parties cannot weigh in on the implementation of AI *ex ante*, then they should at least be able to do *ex post*. One way of accomplishing this would be to take active steps to ensure that potential plaintiffs are made aware of both their right to initiate civil actions and the proper process for doing so in cases where they do have standing. Lastly, we might consider where to draw the line with respect to holding agencies responsible for fostering public awareness. Principle 2 states that "[a]gencies are also encouraged, to the *extent practicable*, to inform the public and promote awareness and widespread availability of standards and the creation of other informative documents" (emphasis added). It is worth asking both why and how agencies should limit their efforts at promoting public awareness with respect to practicality. In certain contexts, it can be exceedingly important for people to know how AI influences them—to the point that failing to help them do so would fly in the face of traditional notions of justice. This is especially relevant in (though not limited to) criminal cases, in which a given defendant's liberty and even life might be at stake.

PRINCIPLE 3 - SCIENTIFIC INTEGRITY & INFORMATION QUALITY

This principle affirms that agency actions on AI must accord with ongoing Federal policies and statutory requirements regarding scientific integrity and information quality. The principles of scientific integrity demand that policy and regulatory decisions are informed by rigorous science, free of political interference. To that end, it is crucial that agencies are expected to hire the diverse expertise necessary to make the best-informed decisions possible on complex questions involving AI.

Commenters would like to re-emphasize that regulation is not necessarily in tension with innovation and may even serve to elevate the public's trust in AI, a primary goal of Executive Order 13859 and the Draft Memorandum. Rigorous and objective scientific study of a given issue may lead an agency to determine that regulation is in fact necessary in at least some cases.

The best practices articulated in Principle 3 should also include an assessment of a system's safety, security, and robustness, the role of human

oversight, the explainability and interpretability of a system, and accountability, monitoring, and liability mechanisms.

Of particular concern is the impact of AI driven systems on protected categories and vulnerable populations. Big data models and machine learning, among other advanced predictive techniques, present a novel challenge to traditional input-based methods of addressing discrimination. As recently discussed in a study by scholars Talia B. Gillis and Jann Spiess, model inputs may correlate with protected characteristics in unanticipated ways, and the formal exclusion of protected characteristics and their close proxies can be insufficient.[14] We urge agencies that are engaged in the evaluation of an AI system's accuracy and fairness to militate against a purely input-oriented evaluative approach and to strongly consider the adoption of an outcome-oriented standard where practicable.[15]

The condition that data must be of sufficient quality for its intended use is too vague. At a minimum, the Draft Memorandum should draw a clear link between this expectation and principles and practices found in the 2020 Federal Data Strategy. Unrepresentative data can increase bias and decrease the overall accuracy of a system—but even a high quality and representative dataset can reflect historical biases in harmful ways. Quality measures for data such as accuracy, consistency, validity, and representativeness are necessary conditions, but may be insufficient in instances where AI is used to make high-stakes decisions such as in criminal justice, lending, or health.

PRINCIPLES 4 AND 5 - RISK ASSESSMENT & MANAGEMENT AND BENEFITS & COSTS

Principles 4 and 5 of the Draft Memorandum address the closely related issues of: (a) risk assessment and risk management; and (b) benefit-cost analysis. Based on our experiences examining public and

---

[14] *See* Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 468-470.

[15] Cathy O'Neil et al., *Regarding Docket No. FR-6111-P-02, HUD's Implementation of the Fair Housing Act's Disparate Impact Standard*, Federal Register, (2019), http://clinic.cyber.harvard.edu/files/2019/10/HUD-Rule-Comment-ONEIL-10-18-2019-FINAL.pdf

private sector organizations developing, deploying, and using AI technologies, we believe there are several issues that the OMB could take into consideration when refining this language.

We commend the OMB for its critical directive noted under Principle 5 that agencies "carefully consider the full societal costs, benefits, and distributional effects before considering regulations related to the development and deployment of AI applications." This language from the OMB highlights two essential issues:

> Range of effects: The use of AI is likely to have a range of societal costs and benefits. Some of these impacts will be obvious and quantifiable, others opaque and indeterminate. Some will be localized, others widespread. Given these distinctions, it is important to consider the full societal impacts of AI—not only the quantifiable, obvious, and widespread ones.
>
> Distribution of effects: The impacts of AI are likely to be unequally distributed around the United States and around the world.[16] Something benefitting urban areas might have costs in rural communities. For example, the deployment of autonomous vehicles that rely on detailed mapping for navigation might produce significant gains in well-mapped cities, but may systematically exclude people in rural areas whose communities are mapped poorly, if at all.

Together, these two points (addressing AI's range and unequal distribution of societal impacts) create important challenges when considering risk assessment and risk management of AI. Because the possible impact will be varied and nuanced, it is important to bear in mind that AI-related risks likely cannot be boiled down to a single metric; what is high risk for one community or population might be low risk for another.

---

[16] For more on the Berkman Klein Center's efforts to examine the strained relationship between artificial intelligence and inclusion, visit https://aiandinclusion.org/. In addition, the International Development Research Center has published a white paper calling for greater caution in using artificial intelligence to avoid contributing to pre-existing forms of inequality and social instability (https://www.idrc.ca/sites/default/files/ai_en.pdf).

And in some cases, the societal impacts of AI might be impossible to quantify or measure at all. For that reason, the OMB should encourage agencies involved in risk assessment to take a broad approach, consulting all possible stakeholders and considering the full societal impacts of AI—even in cases that do not lend themselves to straightforward quantification.

Lastly, it might be helpful to consider narrow areas where regulatory guidance might actually reduce uncertainty and thus enable greater investments and innovation. Given the inherent uncertainties surrounding AI's societal impacts, well-tailored regulation might actually help organizations prioritize responses to the many potential costs and benefits of AI. For that reason, appropriate and narrow regulation should not be dismissed out of hand.

PRINCIPLES 6 AND 10 - FLEXIBILITY AND INTERAGENCY COORDINATION

The Draft Memorandum recommends that agencies adopt "performance-based and flexible approaches that can adapt to rapid changes and updates to AI applications." It notes that regulatory and non-regulatory approaches that rely on rigid, design-based features will be "impractical and ineffective, given the anticipated pace with which AI will evolve and the resulting need for agencies to react to new information and evidence."

The OMB's guidance aims to promote technology-neutral ("tech-neutral") regulatory and non-regulatory approaches to AI. It should note, additionally, that tech-neutral approaches are not merely practical solutions. Tech-neutral approaches also promote flexibility, innovation, and harmonization:[17]

> Flexibility: As the Draft Memorandum notes, performance-based approaches prevent regulatory obsolescence in light of new technologies. Such approaches and conformity assessment schemes promote inter-agency cooperation and give regulators the ability to

[17] Michael Birnhack, *Reverse Engineering Informational Privacy Law,* 15 YALE J. L. & TECH. 24, 27 (2013) (noting the "several important goals" of tech-neutral approaches).

> immediately apply current regulatory and non-regulatory approaches to each and every new technological development.
>
> Innovation: Tech-neutral approaches promote the research and development of new technologies by minimizing legal and regulatory hurdles created by rigid, design-based systems. Developers and funders of new technologies will not be innovating in an unknown regulatory environment; the same existing performance-based approaches and guidance apply.
>
> Harmonization: Performance-based approaches and conformity assessment schemes allow for the streamlined diffusion of newly developed technologies across regulatory jurisdictions.

Furthermore, by adopting and promoting flexible, performance-based approaches domestically, the United States would also promote the adoption of flexible, performance-based international standards, which would ensure "that America remains at the forefront of AI development," and that American AI technologies can quickly diffuse into international markets, fulfilling the Draft Memorandum's goal of reducing barriers to deployment and use of AI technologies.

As Principle 10 notes, a "coherent and whole-of-government approach to AI oversight requires interagency coordination." Such cooperation and interagency information sharing on AI policies would be hindered if such policies were tied to rigid, design-specific regulatory and non-regulatory approaches that could not be generalized. Flexible, performance-based approaches would facilitate interagency coordination as agencies may share and learn from each other's experiences in ensuring "consistency and predictability of AI-related policies that advance American innovation and growth in AI."

PRINCIPLE 7 - FAIRNESS & NON-DISCRIMINATION

The Draft Memorandum recommends that agencies "consider . . . fairness and non-discrimination" "in accordance with law." There are three shortcomings with this approach. First, the OMB does not propose a definition of fairness or non-discrimination in this context, nor does it instruct agencies to cooperate in arriving at a shared definition. There are multiple modes of measuring fairness, and not all are cross-compatible.[18] The OMB should explicitly instruct agencies not only to consider multiple definitions of fairness, but also to work together to insure that fairness is evaluated consistently across government programs. The same is true of non-discrimination. This is best encapsulated in the debate between opportunity and outcome as measures of equality. By failing to address this debate, the OMB guidance does not meaningfully describe the goals of non-discrimination as applied to artificial intelligence.

Second, by including the phrase "in accordance with law," the OMB suggests that agencies should focus on meeting the baseline standard set out by anti-discrimination law. This is a missed opportunity to encourage agencies to go above and beyond the legal minimum and implement best practices that eliminate not just invidious discrimination, as required by the Constitution, but disparate outcomes as well.

Third, the OMB instructs agencies to consider whether adopting AI applications will reduce existing levels of discrimination. While reducing discrimination is an appropriate and, indeed, necessary goal, this language could be interpreted in a problematic manner. Specifically, agencies may interpret this guidance as saying that an AI solution that is less discriminatory than current practices is necessarily legally sufficient to meet the government's burden of not discriminating under the color of law. The OMB should clarify that agencies are expected to provide services in a way that is *not* discriminatory, not merely *less* discriminatory.

---

[18] *See* Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YouTube (March 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk.

PRINCIPLE 8 - DISCLOSURE & TRANSPARENCY

The Draft Memorandum notes that "applications of AI could increase human autonomy," but encourages agencies to "consider the sufficiency of existing or evolving legal, policy, and regulatory environments before contemplating additional measures for disclosure and transparency." The OMB's guidance interprets "appropriate disclosure and transparency" as "context specific," clarifying that the magnitude of potential harms, the technology involved, and the potential benefits of AI applications may all affect transparency and disclosure requirements.

This recommendation assumes that disclosure and transparency are tradeoffs that require sacrificing the benefits of AI. Transparency should not be cast as a "necessary evil" but rather as an essential mode of protecting individual and community rights against dangers posed by AIs. Under the current state of the art, transparency may be accomplished in multiple fashions without impeding performance. For simple AI applications, the underlying statistical model can be made available for public scrutiny. For more complex applications, explainability methods such as explanation-by-counterfactual can be used to provide, if not full transparency, significant insight into the operations of an AI application.[19]

Furthermore, while the OMB is correct that the need for disclosure and transparency is greater where the potential harms are more significant, it is dangerous to suggest that the potential benefits of an AI application could justify use of a non-transparent "black box" technology. Black box technology poses a number of risks, including barriers to recovery for individuals that are harmed by automated decisions made by such systems as well as by unforeseen interactions with such systems.[20] The OMB should clarify that the potential upsides of an AI application do not justify foreclosing avenues of redress for harms caused by that application. At the very least, the OMB should state that non-transparent, undisclosed AI

---

[19] *See, e.g.*, Sandra Wachter et al., *Counterfactual Explanations without Opening the Black Box*, 31 HARV. J.L. & TECH. 841 (2018).
[20] *See, e.g.*, Jonathan Zittrain, *The Hidden Costs of Automatic Thinking*, NEW YORKER (Jul. 23, 2019), https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking

technologies can only be used where separate provisions are made to allow legal compensation for any unintended harms.

PRINCIPLE 9 - SAFETY & SECURITY

The section of the Draft Memorandum titled "Non-Regulatory Approaches to AI" encourages agencies to use their authority to exempt pilot programs, hackathons, and other early-prototype projects from regulation. It goes on to recommend that agencies use early-prototype projects to gather data on the performance of AI systems. This language has concerning implications for safety and security, and for the topic of AI principles as a whole.

As an initial matter, this blanket recommendation does not contain any limiting language, suggesting that even fundamental safeguards on safety and security, fairness and non-discrimination, and disclosure and transparency could be waived. This language, like other passages addressed above, fails to prioritize basic safety and civil liberties over short-term gains in innovation and development. The OMB should, at the very least, revisit this language and clarify that regulations embodying certain governing principles laid out in the Draft Memorandum, including safety, fairness, and disclosure are inviolable and cannot be waived by regulatory agencies.

The idea of exemptions for pilot programs raises special concerns for AI safety and security. Early prototypes are more likely to pose significant safety and security concerns, as they will necessarily involve deployment of untested AI applications. Such applications are more, not less, likely to have security flaws or unsafe features that pose very real threats to the public. Granting waivers for newer, less tested technologies will severely undermine the OMB's own guidance regarding safety and security. The Commenters propose that the OMB adopt additional language in the "Safety and Security" section that encourages agencies to develop regulations, or at least guidance, for creating safe and secure test environments for pilot AI programs. Such an approach would allow innovation and growth to continue without risking the other values set out in the Draft Memorandum.

Conclusion

Commenters thank OMB for the opportunity to offer their perspective on the Draft Memorandum and welcome further opportunities for engagement on these important issues.