BROOKINGS

COMMENTARY

Opportunities and blind spots in the White House's blueprint for an AI Bill of Rights

Nicol Turner Lee and Jack Malamud

December 19, 2022

In October 2022, the White House Office of Science and Technology Policy (OSTP) published ↗ a Blueprint for an AI Bill of Rights ("Blueprint"), which shared a nonbinding roadmap for the responsible use of artificial intelligence (AI). The comprehensive document identified five core principles to guide and govern the effective development and implementation of AI systems with particular attention to the unintended consequences of civil and human rights abuses. While the identification and mitigation of the intended and unintended consequential risks of AI have been widely known for quite some time, how the Blueprint will facilitate the reprimand of such grievances is still undetermined. Further, questions remain on whether the nonbinding document will prompt necessary congressional action to govern this unregulated space.

The Brookings Center for Technology Innovation hosted a conversation with experts from the OSTP, think tanks, and social justice organizations on December 5, 2022, during which they unpacked key aspects of the Blueprint while debating potential limitations and, in some instances, blind spots that may not have been considered in its development. Some thought was given to whether the harms were too broadly defined and if federal agencies had the resources to adhere to the responsible practices and procedures around AI procurement, use, and consistent audits as outlined in the Blueprint.

In terms of progress before and after the Blueprint's release, at least five federal agencies have adopted guidelines for their own responsible use of automated systems. The Department of Defense's (DOD) Ethical Principles for Artificial Intelligence ↗ and the U.S. Agency for International Development's Artificial Intelligence Action Plan ↗ have both implemented some guidance around government use of AI. The Equal Employment Opportunity Commission (EEOC) has also launched its own AI and algorithmic fairness initiative ↗ in their partnership ↗ with the Department of Labor to "reimagine hiring and recruitment practices." Further, the EEOC is collaborating with the Department of Justice to release guidance ↗ on how the use of AI in employment decisions can discriminate against people with disabilities.

A few federal agencies, including the DOD ↗, the Department of Energy ↗, the Department of Veterans Affairs ↗, and the Department of Health and Human Services ↗, have each established their own centers or offices to implement these guidelines. At least a dozen agencies have issued some sort of binding guidance for the use of automated systems in the industries under their jurisdiction, such as the Federal Trade Commission's business guidelines on Using Artificial Intelligence and Algorithms ↗ and the Food and Drug Administration's principles for Good Machine Learning Practice for Medical Device Development ↗. However, the detail and scope of federal agencies' full adherence to these activities still varies in terms of timeline and deliverables. While the mentioned entities are working toward responsible AI use, other federal agencies, including the Consumer Product Safety Commission, has issued a report ↗ with a proposed framework for evaluating potential AI harms.

Part of these varying degrees of progress within the federal government are due to the number of allotted staff who can effectively and expeditiously implement this new rights-based framework. Another related aspect is how far along they each were in the implementation of the previous guidance ↗ around AI from OSTP leaders in the Trump administration that encouraged similar alignments around responsible use.

The Blueprint also relies upon some consistency in the interpretability and protection of civil rights in automated decisions, particularly in the areas of lending, housing, and hiring. Yet, how widely known civil rights laws are followed within digital domains is still unknown, especially given the opacity of the internet and algorithmic applications. Because the factors affecting how algorithmic models operate are often hidden from

consumers ↗, people are normally in the dark about how they function and the potential threats. Recent studies have found ↗ that, although more algorithmic literacy is needed, particularly regarding their use on social media or new platforms, individuals often have little idea about how these algorithms work or what data they use. While the call for increased data privacy and security is also integral to the Blueprint, more work needs to be done to disentangle how to explicitly apply civil rights laws and protections when online malfeasance occurs.

Finally, the current Blueprint looks to the private sector for self-regulatory management, specifically governing and producing products and services from a consumer rights-based approach. However, this expectation may be too ambitious for companies that generally profit from AI's opacity, including the inferential data that is collected from individual users in the absence of demographic information. Similarly, the Blueprint does not provide mandatory, enforceable guidelines, and the lack of an enforcement regime or a central governing body makes self-regulation insufficient to guard against potential harms.

Perhaps, the greatest limitation is the Blueprint's carving out of law enforcement

The use of AI in law enforcement, especially facial recognition, has raised many legal and ethical considerations and created a great deal of risk. The use of AI in law enforcement can reinforce inequality ↗ and disproportionately impact people of color, leading to false arrests and detainment. The Blueprint's developers missed a major opportunity to implement a rights-based framework to automated decision making and make great strides in AI and criminal justice. For now, the Blueprint avoided detailed scrutiny on these fronts by carving out law enforcement—even among federal agencies that use these technologies, including Capitol police, airport security, and customs and border protection officers. Excluding law enforcement may continue the oversurveillance of certain populations, communities, and individuals under the guise of public safety and national security and will not necessarily reduce the history and manifestation of rampant discrimination against people of color and immigrants. If law enforcement were included in the Blueprint provisions and guidance, it could have offered new guardrails and agency for individuals left with little recourse when misidentified and/or scrutinized by existing and emerging AI technologies.

How to make the national Blueprint stick

Some of the panelists from the Brookings December 5 event offered several recommendations worth describing in this blog, starting with the need for a more sectoral approach to AI governance, explicit inclusion of law enforcement in the Blueprint, and the idea of revisiting the nation's existing civil and human rights precedents to ensure their applicability to the online space.

1. Sectoral approaches will be critical to advancement and enforcement. There is no silver bullet for AI governance, which is evident in the comprehensive and singular legislation introduced in the European Union (EU). The EU is currently considering ↗ a new version of the proposed Artificial Intelligence Regulation Act (AI Act) that has faced significant challenges as a single piece of legislation—challenge (https://www.brookings.edu/research/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/) s that a sectoral approach would largely avoid. Maintaining a more flexible approach in the application of the Blueprint's principles would enable innovative solutions tailored toward the various sectors—such as healthcare, housing, and education—where automated systems are used.

2. Law enforcement must be fully part of the national Blueprint guidance. OSTP should consider either offering an addendum to the existing Blueprint to provide guidelines for law enforcement applications or the including relevant agencies under the purview of the federal government. The practices around the use of AI on human subjects, in addition to procurement policies, should be reviewed, and systems adopted ↗ under a "monolithic technology-procurement model" should be re-evaluated, especially since they "rarely [take] constitutional liberty into account," which illustrates the need for including law enforcement in these initial conversations in order to develop clear guidelines for their use of AI.

3. Existing civil rights regimes should be revisited to ensure their applicability to new and emerging digital rules. Historic civil rights laws and statutes have their roots in resistance efforts that pre-date the digital revolution. While, implicitly, laws that govern fair housing, credit, hiring, and education should apply, they can fall under the radar of contemporary interpretation, especially among AI systems that skirt around the edges of civil rights protections. Kearns and Roth pointed (https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/) to the inferential data economy in their signature book,

The Ethical Algorithm (2019), and when left unchecked, demographic data that is extrapolated rather than confirmed by subjects can lead to online biases and disparate impact, completely foreclosing opportunities for historically marginalized populations. The White House should create a new commission on civil rights to evaluate the agility of existing laws to the new digital environment.

Conclusion

A last alternative, but certainly not nominal consideration, is congressional action through constructive legislation that reinforces the insights gleaned from the Blueprint. The work to democratize AI in the U.S. and globally is not an easy task. Unsettled data privacy laws also complicate the presentation of a rights-based framework that potentially gives consumers agency over the data and decisions. But at this point, legislation might be the only way to make these rights stick, given the challenges in enforcing certain criteria and outcomes that may be driven by proprietary interests.

Congress must absolutely look to new data privacy rules that support the implementation of a rights-based AI governance framework, while setting clear guidance around auditing automated decisions within certain use cases—mainly credit, housing, hiring, education, and healthcare. The principles laid out in the Blueprint for an AI Bill of Rights are crucial—individuals must absolutely have the right to safe automated systems, protection from algorithmic discrimination, data privacy, notice of the use of AI, and meaningful human alternatives—but without congressional action, strategies to effectuate change may lack the effective, credible enforcement regime that only legislation can create. Whether Congress will act to codify these principles and expand their coverage to law enforcement and national security, where they are needed most, remains to be seen, but the influence of the White House behind this issue is an encouraging first step.

AUTHORS

Nicol Turner Lee Senior Fellow - Governance Studies, Director - Center for Technology Innovation @drturnerlee

Jack Malamud Research and Administrative Assistant