Blueprint for an AI Bill of Rights

MAKING AUTOMATED SYSTEMS WORK FOR
THE AMERICAN PEOPLE

A mong the great challenges posed to democracy today is the use of technology, data, and automated systems in ways that threaten the rights of the American public. Too often, these tools are used to limit our opportunities and prevent our access to critical resources or services. These problems are well documented. In America and around the world, systems supposed to help with patient care have proven unsafe, ineffective, or biased. Algorithms used in hiring and credit decisions have been found to reflect and reproduce existing unwanted inequities or embed new harmful bias and discrimination. Unchecked social media data collection has been used to threaten people's opportunities, undermine their privacy, or pervasively track their activity—often without their knowledge or consent.

These outcomes are deeply harmful—but they are not inevitable. Automated systems have brought about extraordinary benefits, from technology that helps farmers grow food more efficiently and computers that predict storm paths, to algorithms that can identify diseases in patients. These tools now drive important decisions across sectors, while data is helping to revolutionize global industries. Fueled by the power of American innovation, these tools hold the potential to redefine every part of our society and make life better for everyone.

This important progress must not come at the price of civil rights or democratic values, foundational American principles that President Biden has affirmed as a cornerstone of his Administration. On his first day in office, the President ordered the full Federal government to work to root out inequity, embed fairness in decision-making processes, and affirmatively advance civil rights, equal opportunity, and racial justice in America.[i] The President has spoken forcefully about the urgent challenges posed to democracy today and has regularly

called on people of conscience to act to preserve civil rights—including the right to privacy, which he has called "the basis for so many more rights that we have come to take for granted that are ingrained in the fabric of this country."[ii]

To advance President Biden's vision, the White House Office of Science and Technology Policy has identified five principles that should guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence. The Blueprint for an AI Bill of Rights is a guide for a society that protects all people from these threats—and uses technologies in ways that reinforce our highest values. Responding to the experiences of the American public, and informed by insights from researchers, technologists, advocates, journalists, and policymakers, this framework is accompanied by From Principles to Practice— a handbook for anyone seeking to incorporate these protections into policy and practice, including detailed steps toward actualizing these principles in the technological design process. These principles help provide guidance whenever automated systems can meaningfully impact the public's rights, opportunities, or access to critical needs.

Safe and Effective Systems

Algorithmic Discrimination Protections

Data Privacy

Notice and Explanation

Human Alternatives, Consideration, and Fallback

Applying the Blueprint for an AI Bill of Rights

Download the Blueprint for an AI Bill of Rights

Safe and Effective Systems

You should be protected from unsafe or ineffective systems. Automated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system. Systems should undergo pre-deployment testing, risk identification and mitigation, and ongoing monitoring that demonstrate they are safe and effective based on their intended use, mitigation of unsafe outcomes including those beyond the intended use, and adherence to domain-specific standards. Outcomes of these protective measures should include the possibility of not deploying the system or removing a system from use. Automated systems should not be designed with an intent or reasonably foreseeable possibility of endangering your safety or the safety of your community. They should be designed to proactively protect you from harms stemming from unintended, yet foreseeable, uses or impacts of automated systems. You should be protected from inappropriate or irrelevant data use in the design, development, and deployment of automated systems, and from the compounded harm of its reuse. Independent evaluation and reporting that confirms that the system is safe and effective, including reporting of steps taken to mitigate potential harms, should be performed and the results made public whenever possible.

From Principles to Practice: Safe and Effective Systems

Algorithmic Discrimination Protections

You should not face discrimination by algorithms and systems should be used and designed in an equitable way. Algorithmic discrimination occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections. Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way. This protection should include proactive equity assessments as part of the system design, use of representative data and protection against proxies for demographic features, ensuring accessibility for people with disabilities in design and development, pre-deployment and ongoing disparity testing and mitigation, and clear organizational oversight. Independent evaluation and plain language reporting in the form of an algorithmic impact assessment, including disparity testing results and mitigation information, should be performed and made public whenever possible to confirm these protections.

From Principles to Practice: Algorithmic Discrimination Protections

Data Privacy

You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used. You should be protected from violations of privacy through design choices that ensure such protections are included by default, including ensuring that data collection conforms to reasonable expectations and that only data strictly necessary for the specific context is collected. Designers, developers, and deployers of automated systems should seek your permission and respect your decisions regarding collection, use, access, transfer, and deletion of your data in appropriate ways and to the greatest extent possible; where not possible, alternative privacy by design safeguards should be used. Systems should not employ user experience and design decisions that obfuscate user choice or burden users with defaults that are privacy invasive. Consent should only be used to justify collection of data in cases where it can be appropriately and meaningfully given. Any consent requests should be brief, be understandable in plain language, and give you agency over data collection and the specific context of use; current hard-to-understand notice-and-choice practices for broad uses of data should be changed. Enhanced protections and restrictions for data and inferences related to sensitive domains, including health, work, education, criminal justice, and finance, and for data pertaining to youth should put you first. In sensitive domains, your data and related inferences should only be used for necessary functions, and you should be protected by ethical review and use prohibitions. You and your communities should be free from unchecked surveillance; surveillance technologies should be subject to heightened oversight that includes at least pre-deployment assessment of their potential harms and scope limits to protect privacy and civil liberties. Continuous surveillance and monitoring should not be used in education, work, housing, or in other contexts where the use of such surveillance technologies is likely to limit rights, opportunities, or access. Whenever possible, you should have access to reporting that confirms your data decisions have been respected and provides an assessment of the potential impact of surveillance technologies on your rights, opportunities, or access.

From Principles to Practice: Data Privacy

Notice and Explanation

You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you. Designers, developers, and deployers of automated

systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible. Such notice should be kept up-to-date and people impacted by the system should be notified of significant use case or key functionality changes. You should know how and why an outcome impacting you was determined by an automated system, including when the automated system is not the sole input determining the outcome. Automated systems should provide explanations that are technically valid, meaningful and useful to you and to any operators or others who need to understand the system, and calibrated to the level of risk based on the context. Reporting that includes summary information about these automated systems in plain language and assessments of the clarity and quality of the notice and explanations should be made public whenever possible.

From Principles to Practice: Notice and Explanation

Human Alternatives, Consideration, and Fallback

You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter. You should be able to opt out from automated systems in favor of a human alternative, where appropriate. Appropriateness should be determined based on reasonable expectations in a given context and with a focus on ensuring broad accessibility and protecting the public from especially harmful impacts. In some cases, a human or other alternative may be required by law. You should have access to timely human consideration and remedy by a fallback and escalation process if an automated system fails, it produces an error, or you would like to appeal or contest its impacts on you. Human consideration and fallback should be accessible, equitable, effective, maintained, accompanied by appropriate operator training, and should not impose an unreasonable burden on the public. Automated systems with an intended use within sensitive domains, including, but not limited to, criminal justice, employment, education, and health, should additionally be tailored to the purpose, provide meaningful access for oversight, include training for any people interacting with the system, and incorporate human consideration for adverse or high-risk decisions. Reporting that includes a description of these human governance processes and

assessment of their timeliness, accessibility, outcomes, and effectiveness should be made public whenever possible.

From Principles to Practice: Human Alternatives, Consideration, and Fallback

---

Applying the Blueprint for an AI Bill of Rights

While many of the concerns addressed in this framework derive from the use of AI, the technical capabilities and specific definitions of such systems change with the speed of innovation, and the potential harms of their use occur even with less technologically sophisticated tools.

Thus, this framework uses a two-part test to determine what systems are in scope. **This framework applies to (1) automated systems that (2) have the potential to meaningfully impact the American public's rights, opportunities, or access to critical resources or services.** These Rights, opportunities, and access to critical resources of services should be enjoyed equally and be fully protected, regardless of the changing role that automated systems may play in our lives.

This framework describes protections that should be applied with respect to all automated systems that have the potential to meaningfully impact individuals' or communities' exercise of:

Rights, Opportunities, or Access

Civil rights, civil liberties, and privacy, including freedom of speech, voting, and protections from discrimination, excessive punishment, unlawful surveillance, and violations of privacy and other freedoms in both public and private sector contexts;

Equal opportunities, including equitable access to education, housing, credit, employment, and other programs; or,

Access to critical resources or services, such as healthcare, financial services, safety, social services, non-deceptive information about goods and services, and government benefits.

A list of examples of automated systems for which these principles should be considered is provided in the Appendix. The Technical Companion, which follows, offers supportive guidance for any person or entity that creates, deploys, or oversees automated systems.

Considered together, the five principles and associated practices of the Blueprint for an AI Bill of Rights form an overlapping set of backstops against potential harms. This purposefully overlapping framework, when taken as a whole, forms a blueprint to help protect the public from harm. The measures taken to realize the vision set forward in this framework should be proportionate with the extent and nature of the harm, or risk of harm, to people's rights, opportunities, and access.

---

[i] The Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/

[ii] The White House. *Remarks by President Biden on the Supreme Court Decision to Overturn Roe v. Wade.* Jun. 24, 2022. https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/06/24/remarks-by-president-biden-on-the-supreme-court-decision-to-overturn-roe-v-wade/