BLUEPRINT FOR AN

AI BILL OF RIGHTS

MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE

OCTOBER 2022

THE WHITE HOUSE
WASHINGTON

About this Document

The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People was published by the White House Office of Science and Technology Policy in October 2022. This framework was released one year after OSTP announced the launch of a process to develop "a bill of rights for an AI-powered world." Its release follows a year of public engagement to inform this initiative. The framework is available online at: https://www.whitehouse.gov/ostp/ai-bill-of-rights

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget (OMB) with an annual review and analysis of Federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government.

Legal Disclaimer

The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People is a white paper published by the White House Office of Science and Technology Policy. It is intended to support the development of policies and practices that protect civil rights and promote democratic values in the building, deployment, and governance of automated systems.

The *Blueprint for an AI Bill of Rights* is non-binding and does not constitute U.S. government policy. It does not supersede, modify, or direct an interpretation of any existing statute, regulation, policy, or international instrument. It does not constitute binding guidance for the public or Federal agencies and therefore does not require compliance with the principles described herein. It also is not determinative of what the U.S. government's position will be in any international negotiation. Adoption of these principles may not meet the requirements of existing statutes, regulations, policies, or international instruments, or the requirements of the Federal agencies that enforce them. These principles are not intended to, and do not, prohibit or limit any lawful activity of a government agency, including law enforcement, national security, or intelligence activities.

The appropriate application of the principles set forth in this white paper depends significantly on the context in which automated systems are being utilized. In some circumstances, application of these principles in whole or in part may not be appropriate given the intended use of automated systems to achieve government agency missions. Future sector-specific guidance will likely be necessary and important for guiding the use of automated systems in certain settings such as AI systems used as part of school building security or automated health diagnostic systems.

The *Blueprint for an AI Bill of Rights* recognizes that law enforcement activities require a balancing of equities, for example, between the protection of sensitive law enforcement information and the principle of notice; as such, notice may not be appropriate, or may need to be adjusted to protect sources, methods, and other law enforcement equities. Even in contexts where these principles may not apply in whole or in part, federal departments and agencies remain subject to judicial, privacy, and civil liberties oversight as well as existing policies and safeguards that govern automated systems, including, for example, Executive Order 13960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (December 2020).

This white paper recognizes that national security (which includes certain law enforcement and homeland security activities) and defense activities are of increased sensitivity and interest to our nation's adversaries and are often subject to special requirements, such as those governing classified information and other protected data. Such activities require alternative, compatible safeguards through existing policies that govern automated systems and AI, such as the Department of Defense (DOD) AI Ethical Principles and Responsible AI Implementation Pathway and the Intelligence Community (IC) AI Ethics Principles and Framework. The implementation of these policies to national security and defense activities can be informed by the *Blueprint for an AI Bill of Rights* where feasible.

The *Blueprint for an AI Bill of Rights* is not intended to, and does not, create any legal right, benefit, or defense, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person, nor does it constitute a waiver of sovereign immunity.

Copyright Information

Foreword

Among 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.[1] 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."[2]

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 a technical companion—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.

About this Framework

The Blueprint for an AI Bill of Rights is a set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intel-ligence. Developed through extensive consultation with the American public, these principles are a blueprint for building and deploying automated systems that are aligned with democratic values and protect civil rights, civil liberties, and privacy. The Blueprint for an AI Bill of Rights includes this Foreword, the five principles, notes on Applying the The Blueprint for an AI Bill of Rights, and a Technical Companion that gives concrete steps that can be taken by many kinds of organizations—from governments at all levels to companies of all sizes—to uphold these values. Experts from across the private sector, governments, and international consortia have published principles and frameworks to guide the responsible use of automated systems; this framework provides a national values statement and toolkit that is sector-agnostic to inform building these protections into policy, practice, or the technological design process. Where existing law or policy—such as sector-specific privacy laws and oversight requirements—do not already provide guidance, the Blueprint for an AI Bill of Rights should be used to inform policy decisions.

Listening to the American Public

The White House Office of Science and Technology Policy has led a year-long process to seek and distill input from people across the country—from impacted communities and industry stakeholders to technology developers and other experts across fields and sectors, as well as policymakers throughout the Federal government—on the issue of algorithmic and data-driven harms and potential remedies. Through panel discussions, public listening sessions, meetings, a formal request for information, and input to a publicly accessible and widely-publicized email address, people throughout the United States, public servants across Federal agencies, and members of the international community spoke up about both the promises and potential harms of these technologies, and played a central role in shaping the Blueprint for an AI Bill of Rights. The core messages gleaned from these discussions include that AI has transformative potential to improve Americans' lives, and that preventing the harms of these technologies is both necessary and achievable. The Appendix includes a full list of public engagements.

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.

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.

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.

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.

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.

Definitions for key terms in The Blueprint for an AI Bill of Rights *can be found in Applying the Blueprint for an AI Bill of Rights. Accompanying analysis and tools for actualizing each principle can be found in the* Technical Companion.