

AI and Policy Considerations

Sorelle Friedler



Haverford
COLLEGE

DEPARTMENT OF COMPUTER SCIENCE

AI and the future of our food

By Erin Blakemore
February 28, 2022 at 9:00 a.m. EST



A tractor sprays a soybean field during the spring. (iStock)

[Comment 8](#) [Save](#) [Gift Article](#)

Robots. Drones. Artificial Intelligence.

All three are touted as potential saviors for farmers, and are already being deployed on large farms, where they assist with such tasks as managing crops, milking cows and helping farmers make decisions about their land.

The potential benefits are huge. Increases in farm productivity could help feed the approximately 2.4 billion people around the world who experience food insecurity and malnutrition and revolutionize the way farmers use their land.

That could come at a cost. The analysis points out potential flaws in the agricultural data that fuels AI-powered systems and the possibility that autonomous systems could place productivity over the environment. That could lead to inadvertent errors causing overfertilization, dangerous pesticide use, inappropriate irrigation or erosion, risking crop yields, water supplies and soil. And wide-scale crop failures could exacerbate food insecurity.





REPORT | APRIL 20, 2023



AI in Hiring and Evaluating Workers: What Americans Think

62% believe artificial intelligence will have a major impact on jobholders overall in the next 20 years, but far fewer think it will greatly affect them personally. People are generally wary and uncertain of AI being used in hiring and assessing workers

BY LEE RAINIE, MONICA ANDERSON, COLLEEN MCCLAIN, EMILY A. VOGELS AND RISA GELLES-WATNICK

Would you want to apply for a job that uses AI to help make hiring decisions?

% of U.S. adults who say they would or would not want to apply for a job with an employer that uses artificial intelligence to help in hiring decisions

66% say No

32% say Yes

Americans widely oppose employers using AI to make final hiring decisions, track workers' movements while they work, and analyze their facial expressions

% of U.S. adults who say they ___ employers' use of artificial intelligence for each of the following



Note: Those who did not give an answer are not shown.
Source: Survey of U.S. adults conducted Dec. 12-18, 2022.
"AI in Hiring and Evaluating Workers: What Americans Think"



How can policymakers intervene?

Blueprint for an AI Bill of Rights

THE WHITE HOUSE



Safe and Effective Systems

You should be protected from unsafe or ineffective systems.

Algorithmic Discrimination Protections

You should not face discrimination by algorithms and systems should be used and designed in an equitable way.

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.

Notice and Explanation

You should know when an automated system is being used and understand how and why it contributes to outcomes that impact you.

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.



President Biden

@POTUS

United States government official

Artificial Intelligence has enormous potential to tackle some of our toughest challenges.

But we must address its risks.

That's why last year, we proposed an AI Bill of Rights to ensure that important protections for the American people are built into AI systems from the start.

4:05 PM · Apr 4, 2023 · 3.9M Views



President Biden

@POTUS

United States government official

When it comes to AI, we must both support responsible innovation and ensure appropriate guardrails to protect folks' rights and safety.

Our Administration is committed to that balance, from addressing bias in algorithms – to protecting privacy and combating disinformation.

5:05 PM · Apr 4, 2023 · 2.2M Views

A Technical Companion to the Blueprint for an AI Bill of Rights

1 WHY THIS PRINCIPLE IS IMPORTANT:

This section provides a brief summary of the problems that the principle seeks to address and protect against, including illustrative examples.

2 WHAT SHOULD BE EXPECTED OF AUTOMATED SYSTEMS:

- The expectations for automated systems are meant to serve as a blueprint for the development of additional technical standards and practices that should be tailored for particular sectors and contexts.
- This section outlines practical steps that can be implemented to realize the vision of the Blueprint for an AI Bill of Rights. The expectations laid out often mirror existing practices for technology development, including pre-deployment testing, ongoing monitoring, and governance structures for automated systems, but also go further to address unmet needs for change and offer concrete directions for how those changes can be made.

3 HOW THESE PRINCIPLES CAN MOVE INTO PRACTICE:

This section provides real-life examples of how these guiding principles can become reality, through laws, policies, and practices. It describes practical technical and sociotechnical approaches to protecting rights, opportunities, and access.

AI Policy Recommendations

- Consider what redlines you have for AI use
 - E.g., a ban on affective computing by law enforcement
- Focus on impacts, not technical details
 - Civil Rights
 - Safety and efficacy concerns
 - E.g., agriculture
 - Privacy
 - Other sector-specific concerns (e.g., healthcare, education)
 - Impact on workers: replacement, displacement, humans-in-the-loop
 - Environmental impacts: energy and water usage



AI Policy Recommendations

- **Civil Rights** provisions can prohibit algorithmic discrimination and require impact assessments
- **Duty of Care** provisions can require that systems are tested to make sure they're safe and effective *before* use
- **Privacy protections** that include AI based inferences and destruction of models or other future-protective remedies
- **Transparency** requirements that **notify, explain, and require disclosures** can support accountability
- **Human alternatives** so that people can opt-out, appeal, or receive remedies from AI harms
- **Oversight and enforcement** including via regulation and a private right-of-action





EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

PROPOSED MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND
AGENCIES

FROM: Shalanda D. Young

SUBJECT: Advancing Governance, Innovation, and Risk Management for Agency Use of
Artificial Intelligence

Rights-Impacting AI:⁴¹ AI whose output serves as a basis for decision or action that has a legal, material, or similarly significant effect on an individual's or community's:

1. Civil rights, civil liberties, or privacy, including but not limited to freedom of speech, voting, human autonomy, and protections from discrimination, excessive punishment, and unlawful surveillance;
2. Equal opportunities, including equitable access to education, housing, credit, employment, and other programs where civil rights and equal opportunity protections apply; or
3. Access to critical resources or services, including healthcare, financial services, social services, transportation, non-deceptive information about goods and services, and government benefits or privileges.

Safety-Impacting AI:⁴² AI that has the potential to meaningfully impact the safety of:

1. Human life or well-being, including loss of life, serious injury, bodily harm, biological or chemical harms, occupational hazards, harassment or abuse, or mental health, including both individual and community aspects of these harms;
2. Climate or environment, including irreversible or significant environmental damage;
3. Critical infrastructure, including the critical infrastructure sectors defined in Presidential Policy Directive 21⁴³ and the infrastructure for voting and protecting the integrity of elections; or,
4. Strategic assets or resources, including high-value property, information marked as sensitive or classified by the Federal Government, and intellectual property.



DRAFT FOR PUBLIC REVIEW



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

PROPOSED MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young

SUBJECT: Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence

b. Determining Which Artificial Intelligence Is Presumed to Be Safety-Impacting or Rights-Impacting

All AI within the scope of this section that matches the definitions of “safety-impacting AI” or “rights-impacting AI” as defined in Section 6 must follow the minimum practices in Section 5(c) by the appropriate deadline. Agencies must review each use of AI that they are developing or using to determine whether it matches the definition of safety-impacting or rights-impacting.

assessments, interest rate determinations, or financial systems that apply penalties (e.g., that can garnish wages or withhold tax returns);

- J. Decisions regarding access to, eligibility for, or revocation of government benefits or services; allowing or denying access—through biometrics or other means (e.g., signature matching)—to IT systems for accessing services for benefits; detecting fraud; assigning penalties in the context of government benefits; 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.

ii. **Purposes That Are Presumed to Be Rights-Impacting.** Unless the CAIO determines otherwise, covered AI is presumed to be rights-impacting (and potentially also safety-impacting) and agencies must follow the minimum practices for rights-impacting AI and safety-impacting AI if it is used to control or meaningfully influence the outcomes of any of the following activities or decisions:

- A. Decisions to block, remove, hide, or limit the reach of protected speech;
- B. Law enforcement or surveillance-related risk assessments about individuals, criminal recidivism prediction, offender prediction, predicting perpetrators' identities, victim prediction, crime forecasting, license plate readers, iris matching, facial matching, facial sketching, genetic facial reconstruction, social media monitoring, prison monitoring, forensic analysis, forensic genetics, the conduct of cyber intrusions, physical location-monitoring devices, or decisions related to sentencing, parole, supervised release, probation, bail, pretrial release, or pretrial detention;
- C. Deciding immigration, asylum, or detention status; providing risk assessments about individuals who intend to travel to, or have already entered, the U.S. or its territories; 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); translating official communication to an individual in an immigration, asylum, detention, or border context; or immigration, asylum, or detention-related physical location-monitoring devices.
- D. Detecting or measuring emotions, thought, or deception in humans;
- E. In education, detecting student cheating or plagiarism, influencing admissions processes, monitoring students online or in virtual-reality, projecting student progress or outcomes, recommending disciplinary interventions, determining access to educational resources or programs, determining eligibility for student aid, or facilitating surveillance (whether online or in-person);
- F. Tenant screening or controls, home valuation, mortgage underwriting, or determining access to or terms of home insurance;
- G. Determining the terms and conditions of employment, including pre-employment screening, pay or promotion, performance management, hiring or termination, time-on-task tracking, virtual or augmented reality workplace training programs, or electronic workplace surveillance and management systems;
- H. Decisions regarding medical devices, medical diagnostic tools, clinical diagnosis and determination of treatment, medical or insurance health-risk assessments, drug-addiction risk assessments and associated access systems, suicide or other violence risk assessment, mental-health status detection or prevention, systems that flag patients for interventions, public insurance care-allocation systems, or health-insurance cost and underwriting processes;
- I. Loan-allocation processes, financial-system access determinations, credit scoring, determining who is subject to a financial audit, insurance processes including risk

DRAFT FOR PUBLIC REVIEW



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

PROPOSED MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young

SUBJECT: Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence

c. Minimum Practices for Safety-Impacting and Rights-Impacting Artificial Intelligence

Except as prevented by applicable law and governmentwide guidance, agencies must apply the minimum practices in this section to safety-impacting and rights-impacting AI by August 1, 2024, or **else stop using the AI until it becomes compliant**. Prior to August 1, 2024, agency CAIOs should work with their agencies' relevant officials to bring potentially non-compliant AI into conformity, which may include voluntary requests to third-party vendors to take appropriate action (e.g., via updated documentation or testing measures). To ensure compliance with this requirement, relevant agency officials must use existing mechanisms wherever possible, for example, the Authorization to Operate process. An agency may also request an extension or grant a waiver to this requirement through its CAIO using the processes detailed below.

v. **Additional Minimum Practices for Rights-Impacting AI.**

Starting on August 1, 2024, agencies must follow the above minimum practices for AI that is *either* safety-impacting *or* rights-impacting. In addition, agencies must also follow these minimum practices *before* initiating use of new or existing rights-impacting AI:

A. **Take steps to ensure that the AI will advance equity, dignity, and fairness.**

This should include at least:

1. *Proactively identifying and removing factors contributing to algorithmic discrimination or bias.* Agencies must assess whether their rights-impacting AI materially relies on information about a class protected by Federal nondiscrimination laws in a way that could result in algorithmic discrimination or bias against that protected class. Agencies should also assess whether proxies produce undue influence on their rights-impacting AI. In either case, if the AI's reliance on such information results in unlawful discrimination or harmful bias against protected classes, the agency must cease the use of the information before using the AI for decision-making.
2. *Assessing and mitigating disparate impacts.* Agencies must test their AI to determine whether there are significant disparities in the AI's performance across demographic groups, including in the AI's real-world deployment, and, consistent with applicable law, appropriately address disparities that have the potential to lead to discrimination, cause meaningful harm, or decrease equity, dignity, or fairness. If adequate mitigation of the disparity is not possible, then agencies should not use or integrate the AI tool.
3. *Using representative data.* Agencies should ensure that data used to develop, operate, and assess their AI is adequately representative of the communities who will be affected by the AI, and has been reviewed for improper bias based on the historical and societal context of the data.

B. **Consult and incorporate feedback from affected groups.** To the extent practicable and consistent with applicable law and governmentwide guidance, agencies must consult affected groups, including underserved communities, in the design, development, and use of the AI, and use such feedback to inform agency decision-making regarding the AI. In the event of negative feedback, agencies must consider not deploying the AI or removing the AI from use. Agencies are strongly encouraged to solicit feedback on an ongoing basis from affected groups, such as customers,³² Federal employee groups, and employees' union representatives, particularly after significant modifications to the AI or the conditions or context in which it is used. To carry out such consultations, agencies should take adequate steps to solicit input from the groups affected by the AI, which could include:³³

ONLINE CIVIL RIGHTS ACT



LAWYERS' COMMITTEE FOR
CIVIL RIGHTS
U N D E R L A W

ONLINE

(4) CONSEQUENTIAL ACTION.—The term “consequential action” means an act that is likely to have or contribute to a material effect on the impact of, access to, eligibility for, cost of, terms of, or conditions of any of the following:

- (A) Employment, including hiring, pay, independent contracting, worker management, promotion, termination, and labor relations.
- (B) Education and vocational training, including assessment, proctoring, academic integrity, accreditation, certification, admissions, financial aid, and scholarships.
- (C) Housing and lodging, including rental and short-term housing and lodging, home appraisals, rental subsidies, and publicly-supported housing.
- (D) Essential utilities, including electricity, heat, water, municipal trash or sewage services, internet and telecommunications service, and public transportation.
- (E) Health care, including mental health care, dental, vision, and adoption services.
- (F) Credit, banking, and other financial services.
- (G) Insurance.
- (H) The criminal justice system, immigration enforcement, border control, and child protective services, including risk and threat assessments, bail determinations, sentencing, parole, surveillance, unmanned vehicles and machines, and predictive policing.
- (I) Legal services, including court-appointed counsel services and alternative dispute resolution services.
- (J) Elections, including voting, redistricting, voter eligibility and registration, support or advocacy for a candidate for office, distribution of voting information, election security, and administration.
- (K) Government benefits and services, as well as identity verification, fraud prevention, and assignment of penalties.
- (L) Public accommodations.
- (M) Abnormally dangerous activities.
- (N) Any other service, program, or opportunity which has a comparable legal, material, or similarly significant effect on an individual’s life as determined by the Federal Trade Commission through rules promulgated pursuant to section 553 of title 5, United States Code.

(5) COVERED ALGORITHM.—The term “covered algorithm” means—

(A) —

- (i) a computational process that uses machine learning, natural language processing, artificial intelligence techniques, or other computational processing techniques of similar or greater complexity; or
- (ii) a deterministic computational process derived from a process described in clause (i); and

(B) that, with respect to a consequential action—

- (i) creates or facilitates the creation of a product or information;
- (ii) promotes, recommends, ranks, or otherwise affects the display or delivery of material information;
- (iii) makes a decision; or
- (iv) facilitates human decision making.



ONLINE CIVIL RIGHTS ACT



LAWYERS' COMMITTEE FOR
CIVIL RIGHTS
U N D E R L A W

TITLE II – DUTY OF CARE

SEC. 201. DUTY OF CARE.

- (a) IN GENERAL.—It shall be unlawful for a developer or deployer to offer, license, or use a covered algorithm in a manner that is not safe and effective.
- (b) SAFE.—For purposes of subsection (a), a covered algorithm is safe if—
- (1) the developer or deployer has taken reasonable measures to prevent and/or mitigate harms identified by a pre-deployment evaluation or impact assessment as stipulated in Section 102 (a) and 102 (b);
 - (2) use of the covered algorithm as intended is not likely to result in a violation of this Act; and
 - (3) the developer or deployer evaluates the possibility of not offering, licensing, or using the covered algorithm or removing a covered algorithm from use, and reasonably concludes that—
 - (A) use of the covered algorithm is not likely to result in substantial harm to individuals;
 - (B) the benefits to individuals affected by the covered algorithm likely outweigh the costs to such individuals;
 - (C) individuals can reasonably avoid being affected by the covered algorithm; and
 - (D) use of the covered algorithm is not likely to result in deceptive practices.
- (c) EFFECTIVE.—For purposes of subsection (a), a covered algorithm is effective if the developer or deployer has taken reasonable steps to ensure that—
- (1) the covered algorithm functions at a level that would be considered reasonable performance by a person with ordinary skill in the art;
 - (2) the covered algorithm functions in a manner that is consistent with its expected performance and publicly advertised performance;
 - (3) the covered algorithm functions in a manner that is consistent with any publicly advertised purpose or use; and
 - (4) any data used in the design, development, deployment, or use of covered algorithms is relevant and appropriate to the deployment context and the publicly advertised purpose.
- (d) DECEPTIVE MARKETING OF A PRODUCT OR SERVICE.—A developer or deployer shall not engage in deceptive advertising, marketing, or other public statements regarding a covered algorithm that they develop or deploy.
- (e) OFF-LABEL USE.—It shall be unlawful for a developer to knowingly offer or license a covered algorithm for any use other than an intended use evaluated in the pre-deployment evaluation.
- (f) ABNORMALLY DANGEROUS ACTIVITIES.—It shall be unlawful for a developer or deployer to offer, license, or use a covered algorithm to engage in an abnormally dangerous activity.



Resources

- **AI Bill of Rights**
 - <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- **OMB (Draft) Guidance on AI**
 - <https://ai.gov/input/>
- **Lawyers Committee Online Civil Rights Act**
 - <https://www.lawyerscommittee.org/online-civil-rights-act/>

