

As a part of this plan, HHS will work to ensure alignment with White House Office of Management and Budget Memorandum M-24-10 [_____](#) published March 28, 2024, and in consultation with relevant agencies, to ensure consistency between guidance to federal agencies and to STLTs in operationalizing AI and automated systems in public benefits programs and services.

Objectives

H

improve administration of the programs. While this document references States, Tribes, localities, and territories as a group, HHS does recognize that Tribes are sovereign nations and that as recommendations and policies are developed at a program level (e.g., Tribal, State, or Local) HHS will continue to work with States, Tribes, and Territories to ensure that the recommendations and policies are developed in a way that respects the sovereignty of Tribes and the self-determination of States, Tribes, and Territories.

Efficiency: faster transaction processing and quicker responses to the public including for transactions such as applications, eligibility determinations, benefits distribution, and appeals; a reduction in administrative burdens faced by STLT and

detecting or preventing mental-health issues; flagging patients for interventions; allocating care in the context of public insurance; or controlling health-insurance costs and underwriting;

Emergencies: Choosing to summon first responders to an emergency

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

Sentiment analysis/natural language processing to categorize major themes and trends in unstructured text for customer experience and customer satisfaction surveys, helpdesk tickets, or social media posts referencing a benefit program;

Creation of synthetic data for testing information technology systems; and

AI-

Automation bias: processes or practices that deter agency staff from thinking critically about the outputs of an automated or algorithmic system; for example, policies that require a pace that prohibits meaningful review, or practices that penalize employees for questioning a system's outputs;

Customer alienation: recipient inability to access accurate information; reduction in opportunities for human-led customer support; lack of transparency regarding uses of automated or algorithm-enabled technologies and causes of benefits-related decisions;

Due process: recipient ability to understand and appeal decisions, including access to information about the inputs and the working of the automated or algorithmic system, made by automated or algorithm-enabled technologies;

Unmanaged scope: automation and rapid diffusion of inaccurate, unjust, or unsafe algorithms; lack of human intervention/review in decision-making/insufficient processes to address failures of automation; undue reliance and inability to overrule model outputs;

Privacy risks: implicit coercion to share biometrics or other personally identifiable information, privacy risks associated with use of other public data (e.g., social media postings) in automated and algorithmic systems used in agency decisions and activities;

Agency control of automated or algorithmic systems: over-reliance on private sector vendors and proprietary technology, which prevents knowledge of details on system outcomes and decision-making drivers;

Quality assurance of automated or algorithmic systems: lack of standards, technologies, expertise, processes to assess quality and risk of automated or algorithmic technologies or automated systems and to monitor performance on an ongoing basis; lack of public scrutiny;

Insufficient governance and oversight capabilities and processes: to make informed procurement decisions, manage vendors, and audit systems.

Policy recommendations

HHS highly encourages STLTs to adopt already existing AI best practices as described herein. This plan provides recommendations based on previously issued federal guidance pertinent to the use of AI in public benefits. It also provides a list of potential future guidance and/or recommendations that may be issued at the HHS- or program-level as per below, which should further help clarify practices for STLTs.

Of note, using AI-enabled automated and algorithmic systems does not change programs' responsibility to comply with existing applicable federal, state, and local laws, and regulations, including those addressing privacy, confidentiality, intellectual property, cybersecurity, human and civil rights, and civil liberties.

Recommendations from existing guidance

The federal government has numerous guidance and policy documents pertaining to automated and algorithmic systems. These include the [Blueprint for an AI Bill of Rights](#), the [National Institute of Standards and Technology \(NIST\) AI Risk Management Framework](#), Ex

Ensuring safety and security

The NIST AI Risk Management Framework outlines key elements of safe operations of automated and algorithmic systems:

- responsible design, development, and deployment practices;
- clear information to deployers on responsible use of the system;
- responsible decision-making by deployers and end users;
- explanations and documentation of risks based on empirical evidence of incidents.

Traditional procurement approaches that evaluate systems during the purchasing process but apply little or no diligence to ongoing monitoring will not suffice for automated or algorithm-enabled technologies which are dynamic in nature. Other key elements include the ability to shut down, modify, or have human intervention into systems that deviate from intended or expected functionality. This is especially important in automated systems to prevent rapid proliferation of safety issues, as is establishing mechanisms and processes for identification and reporting of adverse events.

Negative impacts will likely occur without developing responsible AI practices with strong DEIA practices.

Bias can occur when employing AI even in the absence of prejudice, partiality, or discriminatory intent.¹³

These factors include, but are not limited to, tools, datasets, system flows, and business processes. These assessments should not be a one-time activity; they should occur before any procurements and also be integrated into processes for design, development, implementation, testing, training, and ongoing monitoring.¹⁴ Assessments should also consider the risks of AI beyond its intended use.

Rights-impacting and safety-impacting uses of AI should be reevaluated for bias on at least an annual basis and after any significant modification to the AI or the conditions or context in which the AI is used. Periodic reevaluation is important to detect any emergent biases as participant demographics and program rules change over time. If an evaluation determines that inaccurate, biased, or disparate outcomes are produced, the AI use should be discontinued. Evaluations should assess disparate impacts on classes protected by federal nondiscrimination laws and should also assess impacts on underserved communities that are particularly vulnerable to additional burden, barriers, or disruptions in benefit access.

recommend

¹⁵ Each of these biases can occur in the absence of prejudice, partiality, or discriminatory intent.

Systemic bias can be present in datasets, the organizational norms, practices, and processes across the system lifecycle, and the broader society that uses automated or algorithmic systems. For example, if there is racial bias in historical child welfare placements and separation decisions, using longitudinal data to train a decision support tool has a high risk of perpetuating bias. Statistical and computational biases, also known as algorithmic biases, occur when the datasets used to train and update AI models lack adequate representation of different groups or demographics in the data, leading to skewed outcomes. For example, in order for chatbots to answer user questions accurately, they must be trained on input speech from a diverse user base

¹² Executive Order on the Safe and Secure and Trustworthy Development and Use of Artificial Intelligence, October 30, 2023, Section 7.2.

¹³

, NIST Publication 1.09581 0 Td Tw 1c10.9 0.ETdTJ0 Tc5TdTd T

reflecting the full geographic, ethnic, and age range of intended users. AI-enabled program design recommendations must be based on a diverse group of program users, not just those for whom large volumes of existing training material were readily available.

Human biases relate to how an individual or group perceives system information to inform a decision or fill in missing information, or how humans think about purposes and functions of a system. Human-cognitive biases are omnipresent in institutional, group, and individual decision-making processes across the AI lifecycle and system use, including the design, implementation, operation, and maintenance of automated and algorithmic systems. For example, a human may overly trust an AI-generated draft document or decision recommendation as impartial or correct, leading them to insufficiently review and reconsider the AI content.

Preserving options to opt out of AI

Even with the above protections in place, some individuals and groups will not want to use AI-enabled systems.

provides the intended validation of AI outputs and does not result in human actions becoming a “rubber stamp” without the expected scrutiny. Refactoring AI business processes to add human oversight is an insufficient alternative to addressing a root cause of bias or errors in an AI system.

STLT agencies may seek to ensure that staff providing human oversight understand how the AI system functions, what an accurate decision looks like, and how to evaluate a system’s decisions. Staff may benefit from understanding the types of errors their role is meant to detect and have a workload appropriate for providing the expected level of oversight. STLT agencies may seek to ensure staff have the authority to override or alter the decision under review and should be able to escalate patterns of errors they have observed for further analysis and remediation.

Staff who provide human oversight for AI-enabled functions may benefit from training on relevant AI topics and should receive sufficient training in business processes to assess outputs of AI functions or models for accuracy. For example, staff who use an AI-enabled tool to advise them in benefit calculations could be trained in how to calculate benefit amounts without the assistance of the AI tool, so they can evaluate the AI’s recommendation.

STLT agencies may seek to regularly evaluate business processes that interact with AI systems by observing their execution under real-world conditions to determine the effect of automation bias. If AI outcomes that are intended to have human review or that are intended to influence, but not direct, decision-making are being accepted without appropriate scrutiny, STLT agencies may stop using the AI until business processes can be refactored and staff can be retrained.

Protecting the Public

Considering the expanding scale of private data collection utilized by AI, it is imperative that AI-based technologies supporting public benefits programs and services be based on ethical principles that prioritize the protection of recipient privacy and civil liberties.

Data collection should be limited in scope, minimized as much as possible, and determined to be strictly necessary for meeting specific, narrowly identified purposes. Data collected based on these specific, narrowly defined purposes should not be used in a different context without assessing for new privacy risks and implementing appropriate mitigation measures, which may include express consent.

Entities creating, using, or governing automated systems should follow privacy and security best practices designed to ensure data and metadata are not utilized beyond the specific consented use case. Best practices could include using privacy-enhancing cryptography or other types of privacy-enhancing technologies or fine-grained permissions and access control mechanisms, along with conventional system security protocols.¹⁶

AI- and algorithm-based systems should be designed and built with privacy protected by default. Privacy risks should be assessed throughout the development lifecycle, including privacy risks from reidentification, and appropriate technical and policy mitigation measures should be implemented. This includes potential harms to those who are not users of the automated system, but who may be harmed by inferred data, purposeful privacy violations, or community surveillance or other community harms.

Strengthen Governance for Automated and Algorithmic Systems

Promoting innovation and managing associated risks for automated and algorithmic systems requires effective and deliberate governance. In general, the goal of governance is to ensure management of risks stemming from reliance on automated or algorithmic system outputs in public benefits programs to inform, influence, decide, or execute program decisions or actions. Ineffective governance could undermine the efficacy, safety, equitableness, fairness, transparency, accountability, appropriateness, or lawfulness of such decisions or actions.¹⁷

STLTs may benefit from the designation of a central Chief AI Officer role to help lead this effort.

It is important to recognize that broad applicability of automated or algorithm-based technologies requires a governance approach integrated with enterprise risk management strategies and processes alongside other system-wide areas such as cybersecurity, privacy, and regulatory compliance. This will ensure that AI gets the required level of governance attention, is integrated with other system-wide focus areas, and will facilitate operational alignment and organizational efficiencies.

The NIST AI Risk Management Framework identifies 6 key pillars of AI governance:

¹⁶

1. Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively.
2. Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.
3. Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle.
4. Organizational teams are committed to a culture that considers and communicates AI risk.
5. Processes are in place for robust engagement with relevant AI actors.
6. Policies and procedures are in place to address AI risks and benefits arising from third party software and data and other supply chain issues.

Applying the BDC(d)-3.8 (oLI /MCID 87 BDC ov)1.6 (r)-27.1 (of)2.6n]TJ0 Tend

c.8 (t)-1.7 (h)- (a)g, g,

0.881 0 Td[(pa)4.2 (r)a a6 (yi).9 (ng)5.8Dec

348035[(o)Td(p)3(2)(M)C(c)53 (y)6(n)165da2.6(v)0T5.g)T]E.4TjE(fdd)2.1f)0TnT(d)TjEAC50n16(ifa)ta-709-e-8 (d)B(m)5ah(e)h

It is recognized that not all this information may be available for or applicable to each automated or algorithm-based technology, and lack of such information does not have to be disqualifying. However, explicit consideration of these elements and thresholds for each is critical to robust governance of automated and algorithmic systems and system auditing.

to help inform the development and implementation of the guidance, tools, and support discussed in this plan.

Through the coming year, HHS will use a variety of participation and engagement activities to:

- Host ongoing listening sessions with advocates and members of affected and underserved communities to understand potential harms and promote compliance with federal civil rights and privacy laws;

- Engage state, local and community partners, including advisory committees to gather feedback and foster discussion about forthcoming guidance, consistent with applicable law and government-wide guidance;

- Conduct Tribal consultation, as well as Urban Confer through the Indian Health Service;

- Broaden public engagement to ensure the consideration of a range of lived experiences and perspectives in policymaking;

- Host webinars and workshops with benefits and service delivery staff in partnership with select non-governmental entities to enable information sharing;

- Share learnings and feedback received with STLTs for incorporation into the design, development, and use of AI, and to inform agency decision-making regarding automated and algorithmic systems;

- Improve STLT access to non-governmental organization-developed training, educational materials (including playbooks, toolkits, frameworks, and other “best practice”-type content).

APPENDIX A

General End-to-End Benefits Program Workflow

The power and utility of automated and algorithmic systems is that they can be used in different ways across all aspects of public benefits program administration. While there is tremendous heterogeneity in the public benefits programs administered by STLTs, there are general functions that all benefits and services programs typically perform. Automated and algorithm-based technologies are being developed in the marketplace that significantly impact each of these areas. As opposed to the direct administration of public benefits, providing services may have a more indirect relationship with the value chain articulated below, but the principles largely hold.

Below is a flow chart that illustrates some of the general processes involved in benefits programs that may use automated and algorithmic systems:²⁰

Exhibit 1: General end-to-end value chain of public benefits programs

Basic descriptions of each of these functions are as follows:

- management; Overall program policy, program, resource, and operational
- : Identification of and outreach to potential recipients;
- : Processes for recipients to apply for public benefits, including pre-population of forms, submission of new information, identity proofing, and data matching to existing records;
- : Policies and processes used to decide which programs recipients may

Program (LIHEAP) ²⁵		
Temporary Assistance for Needy Families (TANF) ²⁶	The TANF program provides states, territories, and Tribes with flexibility in operating programs designed to help low-income families with children achieve economic self-sufficiency. Block grant recipients use their TANF grants to fund monthly cash assistance payments to low-income families with children, as well as a wide range of services that are designed to address one or more of the program's four broad purposes.	1,99M recipients per month on average (FY 2023) [excludes Separate State Programs and Maintenance of Efforts caseloads]

<p>Administration for Community Living (ACL)²⁷ ACL administers aging and disability programs that provide benefits and services in a range of formats and structures. Included below is a sample of relevant programs.</p>		
ACL programs	Description (including HHS role)	Number of beneficiaries (2023)
Older Americans Act (OAA) Programs ²⁸	<p>ACL administers various programs authorized under the OAA, which provide services and support to older adults to help them remain independent and engaged in their communities. These programs include:</p> <ul style="list-style-type: none"> Nutrition Services: Funding for congregate and home-delivered meals. Supportive Services: Assistance with activities of daily living, transportation, caregiver support, and more. Elder Rights Programs: Legal assistance, elder abuse prevention, and long-term care ombudsman services. Disease Prevention and Health Promotion: Programs promoting healthy aging, chronic disease self-management, falls prevention, and evidence-based health promotion initiatives. 	STLTs transfer funds to other providers via subawards/subcontracts to administer services to beneficiaries. OAA programs can serve up to 12M older adults annually.

Do .

Medicaid³²

Provides health coverage to low-income individuals and families in the United States, including children, pregnant women, elderly adults, and people with disabilities

APPENDIX C

Additional Definitions

F

Underserved communities: The term “underserved communities” refers to populations as well as geographic communities that have been systematically denied the opportunity to participate fully in aspects of economic, social, and civil life, and may include Black, Latino, Indigenous and Native American, Asian American, Native Hawaiian, and Pacific Islander persons and other persons of color; members of religious minorities; women and girls; LGBTQI+ persons; persons with disabilities; persons who live in rural areas; persons who live in United States Territories; persons otherwise adversely affected by persistent poverty or inequality; and individuals who belong to multiple such communities. While not mentioned explicitly in these Executive Orders, underserved communities also include individuals with limited proficiency in English, whether they use spoken language, sign language, or other methods to communicate as defined in Executive Orders 13985, 14020, and 14091

