United States Department of Health & Human Services

Plan for Promoting Responsible Use of Artificial Intelligence in Automated and Algorithmic Systems by State, Local, Tribal, and Territorial Governments in Public Benefit Administration

Introduction

The United States Department of Health and Human Services (HHS) funds billions of dollars in public benefits[1] across numerous programs annually. Many of these federally funded programs are administered by state, Tribal, local, and/or territorial government entities (STLTs), serving millions of Americans.

Recent advances in the availability of powerful artificial intelligence (AI)[2] in automated or algorithmic systems (referred to as "automated systems"[3]) open up significant opportunities to enhance public benefits program administration to better meet the needs of recipients and to improve the efficiency and effectiveness of those programs. These capabilities do not come without risk, however, and it is incumbent on government agencies to ensure that these technologies are deployed responsibly to get the best of what they offer without imposing significant risks on those administering or being served by these programs.

To promote the equitable administration of public benefits, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (EO 14110) Section 7.2(b) requires the Secretary of Health and Human Services, in consultation with other agencies, to publish a plan addressing the use of automated or algorithmic systems in the implementation by States and localities of public benefits and services administered by the Secretary. This plan fulfills this EO 14110 requirement, and is also in alignment with the principles outlined in the *National Institute of Standards and Technology Artificial Intelligence Risk Management Framework*, and the *Blueprint for an AI Bill of Rights*.

As called for in EO 14110, this plan addresses the use of AI-enabled automated or algorithmic systems in the implementation by States and localities of public benefits and services funded by HHS, *"such as to promote:*
- *assessment of access to benefits by qualified recipients;*
- *notice to recipients about the presence of such systems;*
- *regular evaluation to detect unjust denials;*
- *processes to retain appropriate levels of discretion of expert agency staff;*
- *processes to appeal denials to human reviewers;*
- *analysis of whether algorithmic systems in use by benefit programs achieve equitable and just outcomes."*

---

[1] HHS recognizes that there are different statutory definitions of "federal public benefit," "public benefit," and "public services." HHS also recognizes that federal agencies and components have used these terms differently than HHS usage in this plan. We note that this plan is a recommendation to STLTs that administer public benefits programs, that entities other than STLTs are not the direct audience for the guidance in this plan (although they may find this guidance useful), and that this plan is not intended to be exhaustive or comprehensive of responsibilities, obligations, and best practices that flow from use of artificial intelligence across all HHS-funded programs.

[2] See Appendix C

[3] Automated or algorithmic systems, or "automated systems", has the meaning set forth in the *Blueprint for an AI Bill of Rights* and refers to "any system, software, or process that uses computation as a whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure."

As a part of this plan, HHS will work to ensure alignment with White House Office of Management and Budget Memorandum M-24-10 *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence* 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

HHS's vision is that STLTs will be better positioned to procure, develop, and maintain responsible AI-enabled automated systems within the next 2-3 years.

The objectives of this plan are to:

1) Provide recommendations to STLTs and their technology vendors for how they should choose, procure, design, govern, and manage AI in the administration of public benefits and services

2) Outline HHS's plans over the next twelve months to offer STLTs support in developing their own policies and practices for using AI in automated and algorithmic systems for public benefits programs and services[4]

Scope

This plan provides recommendations for balancing opportunities and risks in using automated and algorithmic systems in the administration of HHS-funded public benefits programs and services. This plan directly applies only to the use of AI for automated and algorithmic systems while also providing best practices for consideration for all automated and algorithmic systems. The recommendations in this plan provide a foundation for STLT programs to create their own guidance tailored to program needs and operations and in line with their authorities. Although these recommendations are not intended for public benefits and services that are administered in a mixed format between government entities and community-based organizations, STLTs may find them instructive for such programs and HHS encourages mixed format benefits and services programs to align with the recommendations described herein as they find them applicable. HHS intends to initiate Tribal consultation in the coming months.

While this framework does not indicate any change in HHS's approach to overseeing STLT agencies' use of automated systems that do not use AI, STLT agencies are strongly encouraged to apply these principles and best practices where applicable to non-AI enabled automated systems. STLT agencies are particularly encouraged to identify non-AI enabled automated systems with rights-impacting or safety-impacting uses as described below, evaluate those systems for bias or risks to rights or safety, and mitigate potential risks or harms. All STLT systems used to administer HHS programs should protect rights and safety, advance equity, uphold accountability, and engender public trust, no matter what technology is used. HHS welcomes STLT agencies as well as beneficiaries and their advocates, to proactively engage with us to assess how to best support implementation of these principles and best practices, including with respect to automated systems that do not use AI.

The recommendations laid out in this plan are not mandatory and are general in nature. We support STLTs incorporating these recommendations as appropriate. HHS encourages responsible innovation by the STLT agencies that administer public benefits programs and exploration and use of technologies that

---

[4] HHS acknowledges that the ability for STLTs to fully adopt the direction and practices outlined herein will vary based on numerous factors, e.g., internal capabilities, budget, data quality, policy priorities, etc.

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-specific level, HHS will differentiate between those entities as appropriate.

This plan does not supersede existing relevant federal policies or requirements including those that apply to automated and algorithmic systems such as policies that relate to enterprise risk management, privacy, cybersecurity, and civil rights.

HHS recognizes that some STLTs would have to undertake substantial changes in order to meet the vision laid out in this plan. Over the next twelve months, HHS will engage with STLTs directly to understand how it may support STLTs that want to follow HHS recommendations but face resource or other constraints. Based on what the Department learns, HHS plans to develop guides, playbooks, and other supports to help STLTs implement this plan.

AI used in benefits administration

HHS, in alignment with OMB Memorandum M-24-10, is committed to strengthening governance, advancing responsible innovation, and managing risks in the use of AI-enabled automated or algorithmic systems. HHS recommends that STLTs use AI-enabled automated and algorithmic systems to improve program effectiveness and efficiency, educate STLT staff, enhance benefits access, fairness, and customer experience for the recipients they serve. As noted in Executive Order 14110:

> *"Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security."[5]*

AI can be applied in public benefits in both user-facing applications or processes to improve the experience of applicants and benefits recipients; and behind-the-scenes to streamline administrative and operational functions. While AI has the potential to solve some of the key challenges that STLTs face in administering benefits and services, without proper governance and guardrails, such systems can also create significant harms.

Key areas of opportunity

Areas that HHS intends to ensure that STLT partners can capitalize on, and staff and recipients of Federal financial assistance can benefit from, include:

- **Access**: ability to make programs more accessible to underserved communities[6] who currently face barriers to enrolling in and using public benefits;

---

[5] Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
[6] See Appendix C for definition.

- **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 service provider staff;
- **Effectiveness**: more accurate and timely compliance assessments and eligibility determination; advanced analytics to enhance program and benefit design; targeting staff-time to value-added, recipient-focused activities; enhanced program integrity and risk monitoring;
- **Customer experience**: more comprehensive understanding of customer feedback; needs-based personalization of recommendations and customer support; real-time multilingual support; more convenient user authentication and access options (biometrics for program resource access, etc.); timely decision-making on applications and eligibility; more rapid and accurate answers to questions; proactive information on potential program or eligibility changes such as benefits cliffs, etc.

AI and automation technologies are rapidly maturing and diffusing in information technology systems. This will increase the breadth and depth of use cases to which the capabilities of AI can be applied and their potential impacts on public benefits delivery.

Key areas of risk

A key goal of this plan is to catalyze proactive adoption of AI-enabled automated or algorithmic systems-based technologies in use cases where they can improve public benefits delivery for recipients. Balanced against these opportunities, HHS recognizes the importance of deliberate governance and oversight to mitigate the potential and known risks posed by use of automated or algorithmic systems. As stated in EO 14110, "*harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks*."[7]

OMB Memorandum M-24-10 notes that use of AI comes with a wide range of risks, some of which will already be addressed by existing requirements and recommendations. However, AI-based technologies present specific and unique risks to the rights and safety of the public which must be specifically addressed. These include risks related to efficacy, safety, equity, fairness, privacy, transparency, accountability, reliability, appropriateness, or lawfulness of a decision or action resulting from the use of an automated or algorithmic system to inform, influence, decide, or execute that decision or action.

Uses of AI Presumed to Impact Rights or Safety

This framework centers protection of civil rights in the use of AI stemming from its assignment in Section 7.2 of Executive Order 14110, "Protecting Civil Rights Related to Government Benefits and Programs." Section 7.2(a) directs Federal agencies to "use their respective civil rights and civil liberties offices and authorities – as appropriate and consistent with applicable law – to prevent and address unlawful discrimination and other harms that result from uses of AI in Federal Government programs and benefits administration."

AI holds enormous potential to streamline access to public benefits, reduce administrative burden, and improve the customer experience. That potential must be realized on a foundation of preserving the rights and safety for *all* populations affected by a program. The public should not need to trade or risk their

---

[7] Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

rights or safety to benefit from AI, and AI must not be used in ways that exacerbate bias or inequality. AI uses must comply with nondiscrimination laws, rules, and regulations, as applicable.

HHS will work with STLT agencies to manage risks proportionate to the risk presented by a use of AI. This section describes three tiers of risk, from uses that present the highest risk to uses that present the lowest risk. These tiers are not meant to be comprehensive but provide examples of risk categorization.

Uses of AI that are rights-impacting or safety-impacting present the most risk and may require HHS notification, review, and governance. Uses of AI considered rights-impacting or safety-impacting include both AI systems that directly control outcomes and AI systems that influence outcomes or human decision making. For example, an AI tool that directly makes and enacts program eligibility decisions (if allowed under applicable law) would be considered a rights-impacting use of AI, as would an AI tool that recommends eligibility decisions for a case worker to review and act on.

The following table lists specific uses of AI that are presumed to be rights-impacting and/or safety-impacting in OMB Memorandum M-24-10 and provides examples of potential uses of AI in public benefits administration that would, by extension, be considered rights-impacting or safety-impacting.[8] Please note that not all examples are applicable to all HHS programs, as individual programs may have laws prohibiting such uses.

HHS programs may designate other uses of AI as presumed to be rights-impacting or safety-impacting. These uses may or may not mirror uses of AI enumerated as presumed to be rights-impacting or safety-impacting in OMB Memorandum M-24-10, as its lists are not to be considered exhaustive.

Purpose Presumed to Impact Rights or Safety	Program Use Case
Rights-impacting – Appendix I(2)(i) Determining the terms or conditions of employment, including pre-employment screening, reasonable accommodation, pay or promotion, performance management, hiring or termination, or recommending disciplinary action; performing time-on-task tracking; or conducting workplace surveillance or automated personnel management	<u>Workforce management</u> – Assignment of cases to case workers based on complexity analysis. <u>Education and training evaluation</u> – Assessing program participants for suitability for education and training opportunities, matching participants to work opportunities.
Safety-impacting – Appendix I(1)(j) *Rights-impacting – Appendix I(2)(j)* Carrying out the medically relevant functions of medical devices; providing medical diagnoses; determining medical treatments; providing medical or insurance health-risk assessments; providing drug-addiction risk assessments or determining access to medication; conducting risk assessments for suicide or other violence;	<u>Health screening or risk assessment</u> – assessment or screening for referral to additional services or for interventions (e.g., mental health services, physical therapy, occupational therapy, domestic violence support) <u>Safety</u>: Detecting the presence of dangerous weapons or a violent act

---

[8] For the list of purposes for which AI is presumed to be rights-impacting and/or safety-impacting, see Appendix I of OMB M-24-10, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence* (March 28, 2024), https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

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;	<u>Emergencies</u>: Choosing to summon first responders to an emergency
Rights-impacting – Appendix I(2)(l) Making decisions regarding access to, eligibility for, or revocation of critical government resources or services; allowing or denying access—through biometrics or other means (e.g., signature matching)—to IT systems for accessing services for benefits; detecting fraudulent use or attempted use of government services; assigning penalties in the context of government benefits	<u>Benefit administration</u> – Determining eligibility for a benefit at initial application or redetermination, determining benefit amounts, processing changes that affect eligibility or benefit levels, terminating benefits <u>Integrity and enforcement</u> – Analyzing utilization records for program violations or fraud, assessing sanctions or penalties to providers or individuals, authorizing, or disqualifying vendors, markets, or service providers <u>Access to benefit systems</u> – Allowing or denying access to online applications, participant portals, or other systems for accessing benefit services <u>Making caregiving risk assessments</u> – providing recommendations about adoption matching, child protective actions, child custody, protective actions for older adults or disabled individuals
Rights-impacting – Appendix I(2)(m) Translating between languages for the purpose of official communication to an individual where the responses are legally binding; providing live language interpretation or translation, without a qualified interpreter or translator present, for an interaction that directly informs an agency decision or action; effective communication and auxiliary aids and services for persons with disabilities (such as non-AI generated captions or ASL interpretation for deaf individuals)	<u>Translation of program materials</u> – Translation of legally binding materials without validation by and accountability of a human translator, including online and/or paper applications and application guides <u>Live translation</u> – Live translation without a competent interpreter or translator present for an interaction that directly informs an agency decision or action, such as an eligibility interview, mental health screening, appeal, or administrative hearing

In addition to the specific uses described above, the application of AI-enabled automated or algorithmic technologies to the following functions are also generally presumed to be rights- or safety-impacting:

- <u>Program design and management</u>: triaging cases before assigning to case workers, balancing caseload across workers; 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.
- <u>Application processing</u>: processing application inputs for accuracy and completeness; answering certain questions regarding program requirements and applicability.
- <u>Benefits administration</u>: case management including issuing benefits, renewing benefits, issuing notices, managing appeals, providing customer service, health or nutrition screening, 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.

- <u>Benefits change or disenrollment</u>: processing changes on a case that affect eligibility or benefit levels, terminating benefits regarding access to, eligibility for, or revocation of government benefits or services.

This is not intended to be an exhaustive list of rights- or safety-impacting functions or use cases. Additional functions or use cases may require consideration according to specific program and local context.

As described in the policy recommendations below, STLT agencies may have specific responsibilities to identify and notify and/or obtain approval for AI uses that are presumptively rights- or safety-impacting if those AI uses fall under an existing requirement to notify and/or obtain HHS approval.

Uses of AI that may impact rights or safety

Some uses of AI are not presumptively rights-impacting or safety-impacting but could impact rights or safety depending on how they are applied. These uses of AI do not patently influence outcomes or decision-making but do *inform* outcomes or decision-making in ways that could impact rights or safety.

Examples include, but are not limited to:
- <u>Staff support tools</u> – Internal AI-powered chatbots or policy summarization tools used by call center staff, eligibility workers, or other program administration staff to answer questions in real time. These tools may be rights-impacting if they lead to increased rates of incorrect information being provided or adverse outcomes.
- <u>Customer service tools</u> – Public-facing AI-powered chatbots or phone support may become rights-impacting depending on how they are used to support service delivery. HHS programs will work with STLT agencies to consider the impact of risks from the use of AI, such as increasing errors or adverse outcomes, leading eligible populations to believe they are ineligible for public benefits, providing incorrect program information, or creating barriers to accessing public benefits.

Uses of AI that are unlikely to impact rights or safety

Uses of AI-enabled technologies that are not directly or presumptively rights-impacting or safety-impacting and do not patently influence outcomes present lower risks. Examples of these uses include:
- Interactive Voice Recognition (IVR) technology for call centers that uses voice recognition to assist callers in navigating menus and routing calls;
- Optical Character Recognition (OCR) to transcribe information from uploaded documents or paper forms;
- Chatbots using natural language processing to better understand user questions, with human-coded, logic-based preset outputs, not generative AI responses;

- 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-enabled search tools that answer questions about program requirements or policies by directing caseworkers to the relevant section of an official policy manual or other primary source;
- Internal facing chatbots that answer questions about program requirements or policies by directing caseworkers to the relevant section of an official policy manual or other document, where caseworkers are required to cross-check the manual to confirm the citation before proceeding;
- Converting scans of physical documents into machine-readable formats for further analysis and/or improved accessibility;
- Summarizing case management files, relevant research documents, or other large sources of narrative information to provide a starting point for someone to understand high-level situation before diving into details;
- Creating initial drafts of documents or suggesting ways to rewrite a document in plainer language or with better visual layout (similar to what HHS is doing with the simpler Notice of Funding Opportunities project));
- Transcribing conversations as a back-up set of notes;
- Creation of first drafts of translated materials that a human with language expertise will review and approve.

This is not intended to be an exhaustive list. Whether the use of these or any other such technologies are actually rights- or safety-impacting will depend on the specific circumstances and context implemented by the STLT agency.

As described in the policy recommendations below, STLT agencies may have specific responsibilities to identify and notify and/or obtain approval for AI uses that may be rights- or safety-impacting if those AI uses are determined to be rights- or safety-impacting and fall under an existing requirement to notify and/or obtain HHS approval.

Other risks, applicable across risk tiers

The following is a non-exhaustive group of general risks involved with using AI in public benefits activities in addition to rights- or safety-impacting uses. STLTs should consider these risks and potential mitigation options for each new AI application:

- Inappropriate or invalid use: automated or algorithmic-based technologies applied outside of original model design parameters or use of automated or algorithmic-based technologies with low performance such as to reduce or limit benefits access inappropriately (e.g., high rates of procedural terminations and/or erroneous auto-terminations);
- Unintended consequences of otherwise appropriate use: an 'appropriate' use of AI to optimize efficiencies in one area (such as financial metrics) might have a negative impact on service delivery (such as providing positive health care outcomes) elsewhere. In other words, competing priorities or metrics where an AI system is meant to choose one over the other;
- Inattention to equity: discriminatory outputs based on data imbued with underlying and/or systemic inequities and biases; embedded discriminatory profiling of individuals; automation and rapid proliferation of discriminatory outputs;

- 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, Executive Order 14110, and OMB Memorandum M-24-10. By issuing recommendations and by creating additional resources, HHS seeks to help STLTs find ways to align with them as well.

HHS encourages STLTs to adopt the following priorities in their use of AI in automated or algorithmic systems, consistent with OMB Memorandum M-24-10:

1. Advance Responsible Innovation for Automated and Algorithmic Systems

2. Manage Risks from the Use of Automated and Algorithmic Systems
3. Strengthen Governance for Automated and Algorithmic Systems

The above areas, policies, and principles are reflected below in HHS's recommendations to STLTs.

Advance Responsible AI Innovation

HHS encourages STLT administrators of public benefits that HHS oversees to responsibly use AI-enabled automated and algorithm-based technologies to advance the quality of public benefits programs for recipients. For those STLTs that either are already using AI-enabled automated or algorithmic systems or plan to in the near future, HHS will provide recommendations and technical assistance over the next 12 months to *establish and execute a plan to increase their capacity to appropriately utilize, and evaluate, to root out and mitigate risks to rights, safety and equity of AI in automated and algorithmic systems to achieve their missions and take steps to advance the responsible use of AI in automated and algorithmic systems across the public benefits administration enterprise, including across the various functions described above.*

Key enablers of effective adoption of AI in automated and algorithmic systems include:
- IT infrastructure;
- High quality data for use in training, testing, and operating automated or algorithmic systems;
- Appropriate safeguards (e.g., cybersecurity, privacy, confidentiality, and civil rights)
- Technical talent and staff training;
- Sharing and collaboration in areas such as: models and code, data sets, policies, best practices, and lessons learned, technical resources, procurement, and implementation resources, etc.;
- Quality assurance policies and processes to build and measure trust in adopted automated and algorithmic systems;

Manage Risks from the Use of AI

STLT agencies are responsible for notifying HHS and/or obtaining HHS approval for the AI uses outlined in OMB Memorandum M-24-10 if those AI uses fall under an existing requirement to notify and/or obtain HHS approval. STLT agencies must determine which program system changes and functions incorporating AI trigger notice to and/or approval by HHS per existing HHS program requirements, particularly uses that are presumptively or may be rights-impacting or safety-impacting AI.

HHS strongly encourages STLTs using AI-enabled automated and algorithmic systems in administering the programs HHS listed in this plan to seek to mitigate risks stemming from their use of AI. As per OMB Memorandum M-24-10, such risk considerations may include impacts to safety, security, civil rights, civil liberties, privacy, democratic values, human rights, equal opportunities, worker well-being, access to critical resources and services, agency trust and credibility, and market competition. NIST defines the essential building blocks of AI responsibility and trustworthiness to include accuracy, explainability and interpretability, privacy, reliability, robustness, safety, security, and the mitigation of harmful bias.

HHS strongly encourages STLTs to prioritize risk management for AI-enabled automated and algorithmic systems that have the potential to impact the public's rights or safety. HHS strongly encourages STLTs to identify their uses of AI-enabled automated or algorithmic systems which may impact rights or safety through the process described in Section 'Identifying Automated and Algorithmic Systems Which Impacts Rights or Safety'. For rights- and safety-impacting AI, HHS strongly encourages STLT to follow the policy principles described below, as well as the minimum risk management practices in M-24-10 Section 5(c).

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.

HHS strongly encourages STLTs to use a safety-by-design approach across the technology lifecycle for automated and algorithmic systems to maintain operational safety and security, whereby safety considerations are embedded throughout the product lifecycle of AI-enabled technologies. 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.

HHS strongly encourages STLTs to implement the following practices to promote safe development, use, and operations of automated or algorithmic systems:

- Conducting an impact assessment to determine the estimated benefit from the automated or algorithmic system as compared to its potential risks;
- Measuring the quality and appropriateness of the data used in the automated or algorithmic system's training, testing, and prediction;
- Piloting systems in a real-world context before full deployment to assess whether automated or automatic systems are mature enough to be accurate and helpful;
- Leveraging pilots and limited releases with strong monitoring, evaluation, and safeguards in place, to carry out the final stages of testing before a wider release;
- Ensuring human facilitation and intervention where needed in the event that an automated system fails or causes harm without disruption of service delivery. No matter how rigorously an automated system is tested, there will always be situations for which the system fails;
- Conducting ongoing monitoring to track performance of the automated or algorithmic system on a continuous basis, to detect degradations on functionality and to detect changes to impact on rights and safety.

Supporting workers

STLTs must abide by merit staffing requirements and appropriately fulfill collective bargaining obligations where applicable. Deployment of automated and algorithmic systems should empower, not debilitate, or demoralize, staff charged with administering public benefits programs.

HHS strongly encourages STLTs to consult workers and provide adequate training for all staff around developing, using, enhancing, and maintaining automated and algorithmic systems. Automated or algorithm-enabled technologies can eliminate rote or tedious work and use staff resources for higher value tasks leveraging the top of their skill sets. The use of automated or algorithmic systems informed by human domain expertise to support agency experts can enhance the ability of public benefits agencies to

meet the growing complexity and volume of information within staff and resource constraints.[9] However, the utility of automated or algorithmic systems is only as good as the ability of staff to appropriately use and maintain such systems. Training for staff should address the risk of automation bias, which results when system design, organization processes, or culture deter staff from making judgments including overruling or questioning automated or algorithm-based outputs. Program oversight is also paramount; the program owner should have product knowledge and know enough about these systems to understand how AI is being used.

HHS strongly encourages STLTs to ensure system design and policies sustain staff judgment which can be particularly important in complex or nuanced situations. STLTs should consider which decisions or actions could result in a significant impact on rights or safety, and provide additional human oversight, intervention, and accountability in those cases. Additionally, ongoing monitoring of the automated or algorithmic system must include periodic human reviews to determine whether the deployment context, risks, benefits, or agency needs have evolved.

HHS strongly encourages STLTs to seek to exercise control over automated and algorithm-based technologies when needed rather than to hand control over to third party vendors. STLTs should take steps to ensure that their contracts retain sufficient rights to data and any improvements to that data to facilitate STLTs' continued design, development, testing, and operation of automated and algorithmic systems. STLTs should ensure that their contracts afford sufficient information about and access to systems to assess efficacy and rights safety impacts. Additionally, STLTs should consider contracting provisions that protect data inputs and STLT information used by vendors in the development and operation of automated and algorithmic systems, and that protect such data from unauthorized disclosure and use, especially where that data includes personally identifiable information.

Advancing Diversity, Equity, Inclusion, Accessibility, and Civil Rights[10]

HHS has issued a [*Joint Statement on Enforcement of Civil Rights, Fair Competition, Consumer Protection, and Equal Opportunity Laws in Automated Systems*](#) with the Consumer Financial Protection Bureau, Department of Justice, Department of Housing and Urban Development, Department of Education, Department of Labor, Department of Homeland Security, Equal Employment Opportunity Commission, and the Federal Trade Commission. HHS will continue to issue joint guidance with other agencies when appropriate.

HHS strongly encourages STLTs to identify and assess impacts of automated or algorithmic systems on equity and fairness and mitigate algorithmic discrimination when it is present. STLTs should take steps to mitigate discrimination throughout the design, development, implementation, iteration, and ongoing monitoring after deployment. Such an approach could include proactive assessment of equity considerations as part of system design, ensuring model development based on representative and robust data, and ongoing disparity assessment and mitigation.[11] Such assessment should include whether their rights-impacting AI materially relies on information about a class protected by federal nondiscrimination laws, or variables that are proxies for it, in a way that could result in algorithmic discrimination or bias against that protected class.

---

[9] *AI Guide for Government*, https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html
[10] https://www.hhs.gov/ocr/index.html
[11] *Blueprint for an AI Bill of Rights,* https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

HHS, in alignment with Section 7.2 of EO 14110, strongly encourages STLTs to "use their respective civil rights and civil liberties offices and authorities—as appropriate and consistent with applicable law—to prevent and address unlawful discrimination and other harms that result from uses of automated and algorithmic systems in federal government programs and benefits administration."[12]

HHS encourages STLTs to be inclusive of diversity, equity, inclusion, and accessibility (DEIA) throughout the design, development, implementation, iteration, and ongoing monitoring after deployment. 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] *HHS strongly encourages STLTs employing AI to proactively assess and mitigate factors that contribute to bias, algorithmic discrimination, or that create inequitable outcomes for protected classes or underserved communities.* 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.

HHS strongly encourages STLTs to evaluate rights-impacting and safety-impacting uses of AI for biased or inequitable outcomes under conditions that mirror real-world use before they are used in program administration. 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.

HHS recommends STLTs to consider and manage the following three major categories of AI bias identified by NIST: systemic, statistical, and computational, and human.[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] *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST Publication AI 100-1, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[14] General Services Administration, *AI Guide for Government*, https://coe.gsa.gov/coe/ai-guide-for-government/introduction/index.html.

[15] See NIST Special Publication 1270: Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

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. *HHS encourages STLT agencies to provide options to opt-out of the use of public-facing AI for a human alternative, wherever practicable.* Where a human alternative is not feasible, an automated system that does not use AI could be provided. For example, if an applicant contacts a call center during business hours, they could be able to opt out of interacting with an AI-powered virtual agent to talk with a human representative. If the applicant calls after hours, when call center staff are not on duty, they could be able to opt out of the virtual agent to access a multilingual phone tree of common functions, such as checking benefit balances or hearing application status.

STLTs may seek to ensure opt-out options are easy to find and access and be provided in all languages required to be supported. Opt-out options are most useful when broadly accessible to people with disabilities and provide a level of service (e.g., wait time, administrative burden) that is not disproportionate to AI alternatives. When designing opt-out options and services, STLT agencies may consider that people with disabilities; individuals with limited English proficiency; people experiencing disasters, financial shock, or trauma; and people from underserved communities may be more likely to opt out of AI for human alternatives.

Opt-out options may be less critical for back-end processes, employee-facing functions, or for enabling functions, such as scanning forms using AI-enabled OCR. Opt-out alternatives may also be less critical when AI is solely used for the prevention, detection, or investigation of fraud.

Human oversight of automated and algorithmic processes

Uses of AI-enabled technologies that include human oversight or review can present lower risks if appropriate practices for human-in-the-loop processes are followed. In these cases, although the risk is lower when there is sufficient human review of the output, there may be substantial risk if the review does not occur or is done by someone without proper expertise to evaluate the AI output. *Therefore, HHS encourages STLT agencies to establish clear, effective human oversight protocols, document those processes, train staff, and conduct periodic oversight to ensure adherence.* It is also important to ensure that such human oversight does not overcompensate and completely negate the benefits sought from the technology in the first place. HHS programs may seek to review such protocols and oversight mechanisms and require reporting on the use of AI.

Some risks from the use of AI can be mitigated by requiring a "human in the loop", with human oversight of AI functions or requirements that a human approve AI recommendations before they are enacted. It is important to carefully design these "human in the loop" processes to ensure that human oversight

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

HHS strongly encourages STLTs to ensure automated and algorithm-based technologies advance the interests of the recipients of public benefits programs and services and maximize their ability to access public benefits programs. Critical aspects of ensuring that recipients are best served by AI- and algorithm-enabled technologies include:

- Transparency to recipients about the use of such systems, including plain language and accurate multilingual communication about the use of such systems;
- Consulting and incorporating feedback from affected communities and the public;
- Providing the ability to opt out of use of automated and algorithmic systems in favor of an easily accessible human alternative, where applicable, including consideration of the needs of recipients with disabilities or in need of reasonable accommodations;
- Providing mechanisms for the public to provide feedback on the customer experience and to report incorrect information or issues;
- Notifying negatively affected individuals;
- Providing the ability to appeal decisions that were made in reliance on automated or algorithm-based technology outputs.

Safeguarding Privacy, Civil Rights and Civil Liberties

HHS strongly encourages STLTs to work to protect recipients from violations of or undermining trust in privacy protections through design choices that ensure such protections are included by default, including ensuring that only data strictly necessary for the specific context is collected, and with recipient consent.

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]

HHS encourages STLTs to establish governance for a program-wide risk management framework covering use of AI technologies in public benefits programs supported by HHS, with defined roles and responsibilities for clear understanding and communication of AI uses, assumptions, and limitations, and publish on a publicly available website their governance and risk-management framework. 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] *Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.*
[17] OMB Memorandum M-24-10, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*, https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

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 these governance principles to reviewing, approving, and managing automatic and algorithmic-enabled technologies will help ensure a comprehensive and responsive approach to responsible AI innovation.

HHS strongly encourages STLTs to establish formalized processes to explicitly evaluate and put in place risk mitigation steps for automated and algorithm-based technologies used in public benefits decisions and activities for programs supported by HHS.

Development, procurement, and use of automated and algorithm-based technologies should undergo analysis of potential risks and adverse impacts; include practices to mitigate identified risks; and incorporate policies and controls for governance, such as how data is acquired, managed, and used. Automated and algorithm-based technologies should be evaluated on a variety of key elements, including validity, reliability, robustness, fairness, intelligibility, safety, security, and privacy.

HHS strongly encourages STLTs to collect and maintain an inventory of automated and algorithm-based technologies developed or procured and used in public benefits programs decisions and activities including descriptive information regarding the developers, purposes, risks, validation, and maintenance of such technologies. This inventory should be made public to the extent possible.

Sound governance requires visibility and awareness of automated and algorithmic systems being used across the public benefits administration enterprise, and transparency regarding the development and maturity of such technologies. HHS strongly encourages the inventory of automated and algorithmic systems capture key elements for each application including:[18]

1. Developer details and designed output
2. Purpose of the technology
3. Out-of-scope uses
4. Development details and input features
5. Process used to ensure fairness in design
6. Process to ensure integrity and suitability of data for the machine learning task
7. External validation process
8. Quantitative measures of performance in test environments
9. Quantitative and qualitative measures of performance and efficacy in real-world environments

---

[18] See framework included in the factsheet ONC developed for the decision support interventions certification criterion, from the Office of the National Coordinator for Health Information Technology's "Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1)" final rule, available at: https://www.healthit.gov/sites/default/files/page/2023-12/HTI-1_DSI_fact%20sheet_508.pdf. Resources for specific application of this framework to public benefits activities will be provided as part of resources for STLTs.

HHS strongly encourages STLTs to ensure that this information be included by vendors included in procurement requirements of AI-based technologies or by developers for agency-built applications. 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.

HHS strongly encourages STLT agencies to use the acquisition process, if applicable, to gather information from vendors that will assist with identifying and mitigating risks from the use of AI, including but not limited to:

- Intended use – Vendors should identify uses that are suitable for their AI and purposes, contexts, and uses that are not suitable for their AI's use;
- Training, test, and evaluation data – Vendors should provide details about the model's training data, including the source of the data, the time periods the data describes, and the extent to which the data reflects real-world contexts. For training data that reflects populations, vendors should provide summary statistics of the population whose data it was trained on that are relevant to a public benefits context (e.g., race/ethnicity, age, gender, income brackets);
- Testing results – Vendors should provide information about how they have tested their system, what contexts they tested it in, findings from their testing, and changes they made to address any issues discovered during the testing process. Vendors should provide testing results broken out by the population demographics that are relevant to a public benefits context (e.g., race and ethnicity, age, gender, and income brackets);
- Audit results – Vendors should disclose whether the system and data have been assessed for bias and/or other risks, the results of any assessments, and any actions taken to mitigate findings.

HHS strongly encourages STLT agencies to only acquire and use AI if a vendor provides all necessary information to evaluate suitability and risk. Purchasing from federal or STLT acquisition schedules may simplify the process of gathering information if disclosure is required by the schedule's terms and conditions.

HHS strongly encourages STLT agencies to consider acquisitions that enable the agency to try AI technologies before committing to expensive or lengthy contracts. Agencies could consider testing the technology in a controlled, isolated environment known as a sandbox or conducting a limited pilot to evaluate suitability, risks, and gains from using AI.

As AI becomes more commonplace, STLT agencies may anticipate that vendors may respond to proposals with solutions containing AI, even if AI features were not requested or anticipated for that acquisition.

Topics of potential future guidance

The next section delineates existing federal policy that HHS encourages STLTs to align with as they use AI in administration of public benefits programs. In addition, HHS will seek to further clarify general and program specific principles through new guidance and/or recommendations in the coming months. HHS is continuing to assess and consider for additional clarification in the next 12 months:

- Parameters around automation. The role of human intervention in administering rights and safety impacting automated and algorithmic systems;

- Appeal rights. When applicants and participants can appeal benefit determinations to a human reviewer for reconsideration due to the use of automated and algorithmic systems;
- Customer support. When must applicants and participants have access to a human being for customer support;
- Generative AI used to support public facing use cases and backend administrative use cases;
- IT and data readiness needed to use algorithmic systems in particular use cases;
- Success metrics and service delivery benchmarks for system implementations, including application turnaround, eligibility determinations, appeals decisions, equity, cybersecurity, and privacy;
- Guidance on how this plan will apply to Tribes;
- Best practices on how to streamline application/eligibility process across chatbots and other customer-assistance tools, and multilingual translation.

HHS will work with STLTs to refine this list to meet STLT needs and priorities.

Providing resources and other supports to STLTs

HHS recognizes that the teams responsible for delivering benefits and services face significant challenges in their day-to-day work, ranging from budgetary constraints and staffing shortages to out-of-date technology systems. Safe and responsible adoption and use of automated and algorithmic systems by STLTs in the administration of public benefits and services will require significant resources. HHS understands that STLTs may face challenges in adopting new practices outlined here considering resource constraints.

Federal agencies, including HHS and the General Services Administration (GSA), have already developed training and practical technical assistance guides that offer IT implementation best practices. HHS seeks to better connect the field with these existing resources.

Specific areas in which HHS **will explore** how it can provide additional supports and/or take additional actions include, but are not limited to:

- Providing information on available funding mechanisms for STLTs;
- Convening learning collaboratives or other knowledge sharing structures;
- Providing training data sets and/or training environments, such as those that reflect the diversity of those eligible for benefits programs or are specific to a particular program;
- Developing frameworks to support AI quality assurance for benefits programs;
- Connecting STLTs with educational implementation guides.[19]

Public Engagement strategy

To develop this plan in 180 days, HHS engaged with key affected communities for preliminary consultation and feedback including state and local benefits and services employees, human services professional associations, civil rights advocates, disability rights advocates, privacy advocates, technologists, and legal and labor representatives. As HHS develops guidance and resources over the next twelve months, it will conduct further engagement with the public, including underserved communities,

---

[19] The content offered in educational guides, models, frameworks, or reference materials will be for technical assistance and will not be enforceable.

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
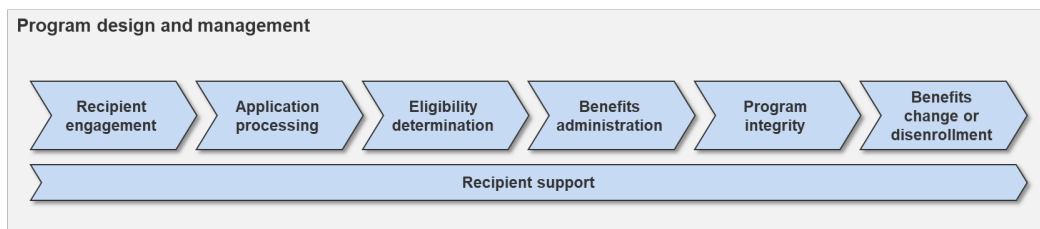
Basic descriptions of each of these functions are as follows:

- *Program design and management:* Overall program policy, program, resource, and operational management;
- *Recipient engagement*: Identification of and outreach to potential recipients;
- *Application processing*: 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;
- *Eligibility determination*: Policies and processes used to decide which programs recipients may be eligible for, including verification of submitted information, calculation of benefit type or amount, risk, or complexity tiering of applications, and preliminary or final determination of eligibility;
- *Benefits administration*: Policies and processes used to administer/distribute program benefits to recipients, including support recipient use of benefits (such as selection of a benefit plan, identification of a provider who accepts a subsidy), and provide updates to recipients about their utilization, manage appeals;
- *Program integrity*: Policies and processes used to confirm recipient eligibility for program benefits received and proper benefit amount, monitor utilization for waste, fraud, and abuse;
- *Benefits change or disenrollment*: Policies and processes used to determine benefits changes based on recipient status changes or policy changes affecting eligibility, and to process renewals;
- *Recipient support*: Direct engagement with recipients to provide assistance throughout the program lifecycle, including customer feedback gathering and analysis;

---

[20] HHS acknowledges that its programs are diverse and not all fit within this structure. This illustration does not limit this plan to programs that this illustration represents.

AI-enabled tools and capabilities are technologies that support decision-making and perform or solve tasks based on algorithms or models that derive relationships from training data and produce outputs used for prediction, classification, recommendation, evaluation, or analysis.[21] In practice, AI technologies are available in many forms and can be applied in public benefits in many ways, such as:

- General purpose tools (e.g., GPT-4, Gemini, Llama);
- Focused tools, such as a fine-tuned versions of Llama or GPT-4, aimed to take into account unique aspects of particular settings or populations;
- User-facing applications, such as chatbots like ChatGPT or interactive voice response (IVR),
- Workflow-specific applications, such as supervised learning algorithms used in task-specific applications (e.g., eligibility determination applications);
- "Front office" tools interacting directly with recipients (e.g., chatbots for customers) or front-line staff in engagements with recipients (bots for real-time support to call center staff);
- "Back office" tools for behind-the-scenes administrative and operational functions (e.g., optical character recognition to transcribe information from scanned materials into databases);
- Traditional AI, which detects patterns, generates insights, and makes predictions;
- Generative AI, which produces newly synthesized content such as narratives or images.

---

[21] For a more formal definition, please see OMB Memorandum M-24-10, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*, https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

APPENDIX B
HHS Public Benefits Programs

HHS supports the administration of numerous public benefits and services[22] that promote the health and well-being of individuals in need. HHS also collaborates closely with other federal agencies in their provision of public benefits and services, including but not limited to the United States Department of Agriculture (USDA), the United States Department of Housing and Urban Development (HUD), and several others.

This plan addresses uses of AI-enabled automated and algorithmic systems in public benefits and services programs that are HHS funded and STLT administered, although these recommendations may be instructive for the equitable use of automated and algorithmic systems in other settings. Public benefits programs include social welfare programs, social insurance programs, tax credits, and other cash, loan, or in-kind assistance programs, particularly those intended to support in-need individuals or communities.

HHS Operating Divisions (OPDIVs) that fund major public benefits programs include the Administration for Children and Families (ACF), Administration for Community Living (ACL), Centers for Medicare & Medicaid Services (CMS), Health Resources and Services Administration (HRSA), Substance Abuse and Mental Health Services Administration (SAMHSA), and others.

Below is a general description of a sample of programs by HHS Operational Divisions that, to the extent that they are carried out by STLTs, may be covered by this plan. As the administration of each program differs in multiple ways, awarding agencies will need to provide specific guidance to grant recipients on the applicability of this plan

Administration for Children and Families (ACF) ACF administers over 60 programs that provide benefits and services in a range of formats and structures. Included below is a sample of relevant programs.		
ACF programs	**Description (including HHS role)**	**~ Number of beneficiaries**
Child Care and Development Fund (CCDF)[23]	CCDF is a block grant to state, territory, and Tribal governments that provides support for children and their families with paying for childcare that will fit their needs and that will prepare children to succeed in school	1.3M children per month on average (FY 2021)
Child support[24]	The federal child support program partners with federal, state, Tribal, and local governments, and others to promote parental responsibility so that children receive reliable support from both of their parents as they grow to adulthood.	12.8M children (2022)
Low Income Home Energy Assistance	LIHEAP assists households with low incomes, particularly those with the lowest incomes that pay a high proportion of household income for home energy, primarily in meeting their immediate home energy needs	5.7M households (2022)

---

[22] HHS recognizes that there are different statutory definitions of "federal public benefit," "public benefit," and "public services." HHS also recognizes that federal agencies and components have used these terms differently than HHS usage in this plan. We note that this plan is a recommendation to STLTs that administer public benefits programs, that entities other than STLTs are not the direct audience for the guidance in this plan (although they may find this guidance useful), and that this plan is not intended to be exhaustive or comprehensive of responsibilities, obligations, and best practices that flow from use of artificial intelligence across all HHS-funded programs.

[23] https://www.acf.hhs.gov/occ/
[24] https://www.acf.hhs.gov/css

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]

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.

ACL programs	Description (including HHS role)	Number of beneficiaries (2023)
Older Americans Act (OAA) Programs[28]	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: • 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.

Centers for Medicare & Medicaid Services (CMS)[29]
CMS regulates and/or funds several health coverage programs through which STLTs administer public benefits. Included below are relevant programs.

CMS programs	Description (including HHS role)	Number of beneficiaries (2023)
Children's Health Insurance Program (CHIP)[30]	Provides health coverage to uninsured children in families with incomes above the Medicaid eligibility threshold but who cannot afford private insurance.	7,180,650 individuals[31]

---

[25] https://www.acf.hhs.gov/ocs/programs/liheap
[26] https://www.acf.hhs.gov/ofa/programs/temporary-assistance-needy-families-tanf
[27] https://acl.gov/
[28] https://acl.gov/about-acl/authorizing-statutes/older-americans-act
[29] https://www.cms.gov/
[30] See generally Title XXI of the Social Security Act, 42 CFR Subchapter D, https://www.medicaid.gov/chip/index.html
[31] https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights/index.html.

Medicaid[32]	Provides health coverage to low-income individuals and families in the United States, including children, pregnant women, elderly adults, and people with disabilities.	77,913,798 individuals
State Marketplace[33]	Provide a way to purchase health insurance coverage for individuals in states that operate their own Marketplace.	17,373,498 individuals

Health Resources and Services Administration (HRSA) [34]
HRSA administers over 90 programs to provide equitable health care to people who are geographically isolated and economically or medically vulnerable. Included below is a sample of relevant programs

HRSA programs	Description (including HHS role)	Number of beneficiaries (2023)
Ryan White HIV/AIDS Program [35](RWHAP)	The RWHAP provides grants to cities, states, counties, and community-based groups. Grants help: Provide care, medication, and essential support services to people with HIV; HIV-related health outcomes; and Reduce HIV transmission. There are five parts of the RWHAP. Part A, B, C and D each have a different purpose and provide grants to STLTs. **Part A**: Eligible Metropolitan Areas (EMAs) and Transitional Grant Areas (TGAs) receive grants to provide medical and support services to cities and counties most severely affected by HIV. **Part B**: All 50 states, District of Columbia, Puerto Rico, U.S. Virgin Islands, and six U.S. territories receive grants to improve the quality of and access to HIV health care and support in the U.S, and provide medications to low-income people with HIV through AIDS Drug Assistance Program **Part C:** Funded local community-based groups provide outpatient ambulatory health services and support for people with HIV and help community-based groups to strengthen their capacity to deliver high-quality HIV care. **Part D:** funded local community-based groups provide medical care for low-income women, infants, children, and youth with HIV, and offer support services for people with HIV and their family members.	In 2022, the RWHAP served more than 550,000 people with HIV.
Maternal, Infant, and Early Childhood Home Visiting (MIECHV) Program[36]	The MIECHV Program supports pregnant people and parents with young children who live in communities that face greater risks and barriers to achieving positive maternal and child health outcomes by supporting the delivery of coordinated and comprehensive high-quality and voluntary early childhood home visiting services to eligible families. Families choose to participate in home visiting programs,	In FY 2023, the MIECHV Program served more than 139,000 parents and children in over 1,000 counties across all 50 states, the District of Columbia, and 5 territories.

---

[32] See generally Title XIX of the Social Security Act, 42 CFR Subchapter C, medicaid.gov
[33] https://www.cms.gov/marketplace/states/state-marketplace-resources
[34] https://www.hrsa.gov/
[35] https://ryanwhite.hrsa.gov/
[36] https://mchb.hrsa.gov/programs-impact/programs/home-visiting/maternal-infant-early-childhood-home-visiting-miechv-program

	and partner with health, social service, and child development professionals to set and achieve goals that improve their health and well-being.	

Substance Abuse and Mental Health Services Administration (SAMHSA)[37]
SAMHSA is the agency within HHS that leads public health efforts to advance the behavioral health of the nation. Included below is a sample of relevant programs=

SAMHSA Programs	Description of the Program	Number of beneficiaries (2023)
State Opioid Response (SOR) Grant	The State Opioid Response (SOR) program addresses the opioid overdose crisis by providing resources to states and territories for increasing access to FDA-approved medications for the treatment of opioid use disorder (MOUD), and for supporting the continuum of prevention, harm reduction, treatment, and recovery support services for opioid use disorder (OUD) and other concurrent substance use disorders.	550,000
Substance Use Prevention, Treatment, and Recovery Services Block Grant (SUBG)[38]	The Substance Use Prevention, Treatment, and Recovery Services Block Grant (SUPTRSBG or SUBG) program allocates funding to all 50 states, the District of Columbia, five (5) US Territories, three (3) Freely Associated States, and the Red Lake Band of Chippewa Indians, to support the provision of substance use disorder (SUD) primary prevention, intervention, treatment, and recovery support services.	1,523,265
Community Mental Health Services Block Grant (MHBG)[39]	The MHBG provides funds to grantees states and U.S. territories to provide comprehensive, community-based mental health services to adults with serious mental illnesses and to children with Serious Emotional Disturbance and to monitor progress in implementing systems through which these services are provided.	8,180,236

---

[37] https://www.samhsa.gov/
[38] https://www.samhsa.gov/grants/block-grants/subg
[39] https://samhsa.gov/grants/block-grants/mhbg

APPENDIX C

Additional Definitions

For the purposes of this plan, the following definitions are applicable:

Artificial Intelligence (AI): The term "artificial intelligence" has the meaning provided in Section 6 of OMB Memorandum M-24-10:

The term "artificial intelligence" has the meaning provided in Section 238(g) of the John S. McCain National Defense Authorization Act for Fiscal Year 2019,[40] which states that "the term 'artificial intelligence' includes the following":

1. *Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.*
2. *An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.*
3. *An artificial system designed to think or act like a human, including cognitive architectures and neural networks.*
4. *A set of techniques, including machine learning, that is designed to approximate a cognitive task.*
5. *An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.*

For the purposes of this memorandum, the following technical context should guide interpretation of the definition above:

1. *This definition of AI encompasses, but is not limited to, the AI technical subfields of machine learning (including deep learning as well as supervised, unsupervised, and semi-supervised approaches), reinforcement learning, transfer learning, and generative AI.*
2. *This definition of AI does not include robotic process automation or other systems whose behavior is defined only by human-defined rules or that learn solely by repeating an observed practice exactly as it was conducted.*
3. *For this definition, no system should be considered too simple to qualify as covered AI due to a lack of technical complexity (e.g., the smaller number of parameters in a model, the type of model, or the amount of data used for training purposes).*
4. *This definition includes systems that are fully autonomous, partially autonomous, and not autonomous, and it includes systems that operate both with and without human oversight.*

Automated Systems: The term "automated system" is any system, software, or process that uses computation as a whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities.

---

[40] Pub. L. No. 115-232, § 238(g), https://www.govinfo.gov/app/details/PLAW-115publ232.

Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure.[41]

Automation Bias: The term "Automation Bias" has the meaning provided in Section 6 of OMB Memorandum M-24-10:

The term "automation bias" refers to the propensity for humans to inordinately favor suggestions from automated decision-making systems and to ignore or fail to seek out contradictory information made without automation.

Equity: The term "equity" has the meaning provided in Section 10(a) of Executive Order 14091.[42]

Generative AI: The term "generative AI" has the meaning provided in Section 3(p) of Executive Order 14110.

Human-In-The-Loop: The term "human-in-the-loop" has the meaning provided in NIST's "The Language of Trustworthy AI: An In-Depth Glossary of Terms":[43]

An AI system that requires human interaction.

Machine Learning: The term "machine learning" has the meaning provided in Section 3(t) of Executive Order 14110.

Public Benefits Program: The term "public benefits program" has the meaning provided in footnote 6 of OMB Memorandum M-22-10:[44]

"[Public benefits programs]" should be construed widely to include social welfare programs; social insurance programs; tax credits; and other cash, loan, or in-kind assistance programs, particularly those intended to support in-need individuals or communities.

Rights-Impacting AI: The term "Rights-Impacting AI" has the meaning provided in Section 6 of OMB Memorandum M-24-10:

The term "rights-impacting AI" refers to AI whose output serves as a principal basis for a decision or action concerning a specific individual or entity that has a legal, material, binding, or similarly significant effect on that individual's or entity'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, insurance, credit, employment, and other programs where civil rights and equal opportunity protections apply; or*

---

[41] The White House. (2023, November 22). *Blueprint for an AI Bill of Rights | OSTP | The White House* https://www.whitehouse.gov/ostp/ai-bill-of-rights/
[42] Executive Order 14091, *Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government*, https://www.govinfo.gov/content/pkg/FR-2023-02-22/pdf/2023-03779.pdf.
[43] NIST, *The Language of Trustworthy AI: An In-Depth Glossary of Terms*, https://airc.nist.gov/AI_RMF_Knowledge_Base/Glossary
[44] OMB M-22-10, *Improving Access to Public Benefits Programs Through the Paperwork Reduction Act* (April 13, 2022), https://www.whitehouse.gov/wp-content/uploads/2022/04/M-22-10.pdf.

3. *Access to or the ability to apply for critical government resources or services, including healthcare, financial services, public housing, social services, transportation, and essential goods and services.*

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

APPENDIX D

Additional resources

The following are additional resources related to AI:

- The NIST AI Risk Management Framework listed above is available at https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. Accompanying the RMF are a companion NIST AI RMF Playbook, AI RMF Explainer Video, AI RMF Roadmap, AI RMF Crosswalk, and various Perspectives.

- The Office of Management and Budget Memorandum M-24-10 "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence" is an additional resource for learning how agencies will be advancing AI governance and innovation while managing risks from the use of AI in the federal government.

- General Services Administration's Artificial Intelligence Community of Practice (AI CoP) serves as a pivotal network for government employees engaged in or interested in artificial intelligence, bringing together over 3,450 members from over 100 federal, state, and local government agencies. The community is dedicated to accelerating the responsible and thoughtful adoption of AI technologies across the government. Membership to GSA's AI CoP is open to all government employees and mission-supporting contractors with a .gov and .mil email address. More information is available at https://coe.gsa.gov/communities/ai.html.

- An inventory of AI use cases by federal agency is available at https://ai.gov/ai-use-cases/.