Framework to Advance AI Governance and Risk Management in National Security

THE WHITE HOUSE
WASHINGTON

Table of Contents

Overview 1

Scope 1

Pillar I: AI Use Restrictions 2

- Prohibited AI Use Cases 2
- High-Impact AI Use Cases 3
- AI Use Cases Impacting Federal Personnel 4
- Additional AI Use Restrictions 4

Pillar II: Minimum Risk Management Practices for High-Impact and Federal Personnel-Impacting AI Uses 5

- Risk and Impact Assessments and Ensuring Effective Human Oversight 5
- Additional Procedural Safeguards for AI Impacting Federal Personnel 7
- Waivers 7
- Chief AI Officers 8

Pillar III: Cataloguing and Monitoring AI Use 8

- Inventory 8
- Data Management 8
- Oversight and Transparency 9

Pillar IV: Training and Accountability 12

Overview

AI is one of the most powerful technologies of our time and presents a significant opportunity for innovation to advance U.S. national security. Such innovation must be responsible, lawful, and align with democratic values, including human rights, civil rights, civil liberties, privacy, and safety. The performance of AI systems shall be such that U.S. Government personnel can have trust and confidence in using AI systems, and the use of those systems should not undermine the public's faith in U.S. national security institutions. The U.S. Government shall use AI systems and employ force informed by AI systems in a manner that complies with all applicable law and policy, including obligations under International Humanitarian Law, Human Rights Law, and the U.S. Government's existing frameworks for the responsible use of AI.

Scope

The Framework to Advance AI Governance and Risk Management in National Security ("AI Framework") builds on and fulfills the requirements found in Section 4.2 of the National Security Memorandum on Advancing the United States' Leadership in AI, Harnessing AI to Fulfill National Security Objectives, and Fostering the Safety, Security, and Trustworthiness of AI ("AI NSM"), which directs designated Department Heads to issue guidance to their respective components/sub-agencies to advance governance and risk management practices regarding the use of AI as a component of a National Security System (NSS).[1,2] This AI Framework is intended to support and enable the U.S. Government to continue taking active steps to uphold human rights, civil rights, civil liberties, privacy, and safety; ensure that AI is used in a manner consistent with the President's authority as commander-in-chief to decide when to order military operations in the nation's defense; and ensure that military use of AI capabilities is accountable, including through such use during military operations within a responsible human chain of command and control. AI use in military contexts shall adhere to the principles and measures articulated in the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, announced by the United States on November 9, 2023. This AI Framework includes four primary pillars relating to:

1. Identifying prohibited and "high-impact" AI use cases based on risk they pose to national security, international norms, democratic values, human rights, civil rights, civil liberties, privacy, or safety, as well as AI use cases that impact Federal personnel.
2. Creating sufficiently robust minimum-risk management practices for those categories of AI that are identified as high impact, including pre-deployment risk assessments.

[1] In the AI NSM, covered Department Heads include the Secretary of Defense, Director of National Intelligence, Attorney General, Secretary of Homeland Security, Secretary of Energy, Secretary of State, Secretary of Treasury, Secretary of Commerce, and any other Department Head of a covered agency that uses AI as part of a National Security System.

[2] This AI Framework is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

3. Cataloguing and monitoring high-impact AI use.
4. Ensuring effective training and accountability mechanisms.

The AI Framework is complementary to, but does not otherwise replace or modify, OMB Memorandum M-24-10. All AI use by federal agencies shall be governed by either OMB Memorandum M-24-10 and its successor policies or by this AI Framework. This AI Framework covers AI when it is being used as a component of an NSS.[3] This AI Framework applies to both new and existing AI developed, used, or procured by or on behalf of the U.S. Government, and it applies to system functionality that implements or is reliant on AI, rather than to the entirety of an information system that incorporates AI.[4]

Updates to this AI Framework shall be made pursuant to a National Security Council (NSC) Deputies Committee meeting through the process described in National Security Memorandum-2 of February 4, 2021 (Renewing the National Security Council System).

Pillar I: AI Use Restrictions

Prohibited AI Use Cases

Covered agencies shall refrain from using AI in any manner that violates domestic or international law obligations and shall not use AI in a manner or for purposes that pose unacceptable levels of risk. Consistent with these goals, covered agencies are prohibited from using AI with the intent or purpose to:

- Profile, target, or track activities of individuals based solely on their exercise of rights protected under the Constitution and applicable U.S. domestic law, including freedom of expression, association, and assembly rights.
- Unlawfully suppress or burden the right to free speech or right to legal counsel.
- Unlawfully disadvantage an individual based on their ethnicity, national origin, race, sex, gender, gender identity, sexual orientation, disability status, or religion.

[3] As defined in 44 U.S.C. § 3552(b)(6).

[4] This AI Framework does not govern: (i) regulatory actions designed to prescribe law or policy regarding non-agency uses of AI; (ii) evaluations of particular AI-enabled applications when the AI provider is the target or potential target of a regulatory enforcement, law enforcement, or national security action; (iii) development of metrics, methods, and standards to test and measure AI, where such metrics, methods, and standards are for use by the general public or the government as a whole, rather than to test AI for a particular agency application; (iv) use of AI to carry out basic research or applied research, except where the purpose of such research is to develop particular AI applications within the agency; (v) evaluation of a potential vendor, commercial capability, or freely available AI capability that is not otherwise used in agency operations, exclusively for the purpose of making a procurement or acquisition decision; or (vi) use of AI in controlled testing conditions to carry out the minimum testing requirements.

- Detect, measure, or infer an individual's emotional state from data acquired about that person, except for a lawful and justified reason such as for the purposes of supporting the health of consenting U.S. Government personnel.
- Infer or determine, relying solely on biometrics data, a person's religious, ethnic, racial, sexual orientation, disability status, gender identity, or political identity.
- Determine collateral damage and casualty estimations, including identifying the presence of noncombatants, prior to kinetic action without (1) rigorous testing and assurance within the AI systems' well-defined uses and across their entire lifecycles, and (2) oversight by trained personnel who are responsible for such estimations exercising appropriate levels of judgment and care.
- Adjudicate or otherwise render a final determination of an individual's immigration classification, including related to refuge or asylum, or other entry or admission into the United States.
- Produce and disseminate reports or intelligence analysis based solely on AI outputs without sufficient warnings that enable the reader of the reports or analysis to recognize that the report or analysis is based solely on AI outputs.
- Remove a human "in the loop" for actions critical to informing and executing decisions by the President to initiate or terminate nuclear weapons employment.

High-Impact AI Use Cases

Some AI activities in military, intelligence, or broader defense contexts that benefit U.S. national security objectives may simultaneously introduce significant new risks. Risks from the use of AI can arise in the event of AI failure, but can also manifest, for example, from ineffective outputs or if the AI was used in a context for which it was not intended. Such high-impact AI activities require sufficient safeguards to mitigate risk. For the purposes of this AI Framework, high-impact AI uses include AI whose output serves as a principal basis for a decision or action that could exacerbate or create significant risks to national security, international norms, democratic values, human rights, civil rights, civil liberties, privacy, or safety, such as the high-impact activities identified in the non-exhaustive list below. Agencies shall review each existing or planned use of AI to determine whether it matches this definition. Consistent with these goals, AI use is presumed to be high impact if the AI use controls or significantly influences the outcomes of any of the following activities:

- Tracking or identifying individuals in real time, based solely on biometrics, for military or law enforcement action.
- Classifying an individual as a known or suspected terrorist, insider threat, or other national security threat in order to inform decisions or actions that could affect their safety, liberty, employment, immigration status, ability to enter or remain in the United States, or Constitutionally-protected rights and freedoms.
- Determining an individual's immigration classification, including related to refuge or asylum, or other entry or admission into the United States.

- The activities referenced in Appendix I of OMB Memorandum M-24-10, when those activities:
    - Occur within the United States;
    - Impact U.S. persons; or
    - Bear on immigration processes or other entry or admission into the United States.
- Designing, developing, testing, managing, or decommissioning sensitive chemical or biological, radiological, or nuclear materials, devices, and/or systems (including chemical and biological data sources) that could be at risk of being unintentionally weaponizable.
- Operating or deploying malicious software designed to allow AI to automatically and without human oversight write or rewrite code in a way that risks unintended performance or operation, spread autonomously, or cause physical damage to or disruption of critical infrastructure.
- Using AI as a sole means of producing and disseminating finished intelligence analysis.

AI used in autonomous or semi-autonomous weapon systems are covered by the policies articulated in Department of Defense Directive 3000.09 and successor or related policies.

AI Use Cases Impacting Federal Personnel

For the purposes of this AI Framework, Federal personnel-impacting AI use cases include AI whose output serves as a significant basis for a decision or action resulting in a legal, material, binding, or similarly significant effect on individual military service members, individuals in the Federal civilian workforce, or individuals to whom the agency has extended an offer of employment. Agencies shall review each existing or planned use of AI to determine whether it matches this definition. Consistent with these goals, AI is automatically presumed to impact Federal personnel if it is used to control or significantly influence the outcomes of any of the following activities:

- Making hiring decisions, including determining pay or benefits packages;
- Determining whether to promote, demote, or terminate employees; and
- Determining job performance, physical health, or mental health diagnoses or outcomes for U.S. Government personnel.

Additional AI Use Restrictions

Department Heads shall, as needed, add prohibited, high-impact, or Federal personnel-impacting categories of AI – applicable to their components' missions, authorities, and responsibilities – to these lists.[5] Department Heads shall maintain unclassified public lists of prohibited and high-impact AI categories they have added to these lists, as well as additional categories they have created for additional oversight and safeguards, but these lists may have classified annexes

[5] In determining these categories of AI, Department Heads should consider the categories of AI in OMB Memorandum M-24-10.

as appropriate to protect classified or controlled information. These lists must be provided to the Assistant to the President for National Security Affairs (APNSA).

Pillar II: Minimum Risk Management Practices for High-Impact and Federal Personnel-Impacting AI Uses

Risk and Impact Assessments and Ensuring Effective Human Oversight

High-impact AI use necessitates additional safeguards. The practices in this AI Framework represent a minimum baseline for managing risk from the use of high-impact AI. Covered agencies must, taking into consideration the broad risk factors outlined in section 4.2(c) of the AI NSM, identify additional context-specific risks that are associated with their AI use and address them as appropriate, including by adding minimum risk management practices. All updates to high-impact AI use and associated minimum risk management practices must be provided to the APNSA.[6]

Within 180 days of the issuance of this AI Framework, covered agencies shall begin following these practices before using new or existing high-impact AI:

- Complete an AI risk and impact assessment, including at least identifying the intended purpose for the AI, its expected benefits, and its potential risks. The AI risk and impact assessment should include:
    - The intended purpose for the AI and its expected benefit, supported by metrics or qualitative analysis, as appropriate. The analysis should demonstrate an expected positive outcome, and it should demonstrate that the AI is better suited to accomplish the relevant task as compared to alternative strategies.
    - The potential risks of using the AI, as well as what, if any, additional mitigation measures, beyond these minimum practices, the agency will take to help reduce these risks. Agencies should document and assess the possible failure modes of the AI. The expected benefits of the AI functionality should be considered against

[6] To address these potential risk management gaps, covered agencies are encouraged to promote and incorporate, as appropriate, additional best practices for AI risk management, such as from OMB Memorandum M-24-10, DOD's Responsible AI Strategy and Implementation Pathway, the Artificial Intelligence Ethics Framework for the Intelligence Community, the Blueprint for an AI Bill of Rights, the NIST AI Risk Management Framework, the DHS Safety and Security Guidelines for AI in Critical Infrastructure, applicable international standards, workforce principles and best practices for employers that could be used to mitigate AI's potential harms and maximize its potential benefits established pursuant to Section 6(b)(i) of Executive Order ("E.O.") 14110, or any successors to those documents. Covered agencies are also encouraged to continue developing their own agency-specific practices, as appropriate and consistent with this AI Framework and the principles in E.O. 13960, 14091, and 14110.

its potential risks, and if the benefits do not meaningfully outweigh the unmitigated risks, agencies should not use the AI.

  - The quality and appropriateness of the relevant data. Agencies must assess the quality, to the extent practicable, of the data used in the AI's design, development, training, testing, and operations, and the data's fitness to the AI's intended purpose. If a covered agency cannot access the data used to train, evaluate, and operate a given AI system, it must obtain sufficient descriptive information from the AI or data provider. At a minimum, covered agencies must document to the extent practicable: (i) the general provenance and quality of the data for the AI's intended purpose for commercially-acquired models; (ii) how the data are relevant to the task being automated and are expected to be useful for AI development, testing, and operation; (iii) whether the data are sufficient to address the range of real-world inputs the AI system might encounter; (iv) whether the data come from a reliable source; and (v) how errors caused by the AI will be adequately measured and reasonably limited.

- Test the AI system sufficiently in a realistic context to confirm it will perform as intended, achieve its expected benefits, and that the associated risks will be sufficiently mitigated once deployed, or else the agency should not use the AI system. Agencies are encouraged to leverage pilots and limited releases, with strong monitoring, evaluation, and safeguards in place, to carry out the final stages of testing before a wider release.
- To the extent practicable, and through an independent reviewing authority not directly involved in the system's development, conduct an evaluation specific to the intended purpose and planned deployment of the AI system and make the evaluation available to the AI Governance Board or equivalent. Qualified personnel could include the Chief AI Officer or other agency officials with test and evaluation responsibilities.
- Identify and mitigate factors that may contribute to unlawful discrimination or harmful bias, including through determining whether the AI model results in significant disparities in the model's performance (such as accuracy, precision, and reliability in predicting outcomes) across demographic groups if applicable.
- Develop processes, including training, to mitigate the risk of overreliance on AI systems (such as "automation bias" or other human factor considerations that may result in insufficient human decision-making).
- Train and assess the AI system's operators, who must have, at a minimum, appropriate training on the specific AI use case, product, or service, including its limitations, risks, and expected modes of failure, as well as a general knowledge of how the AI system functions in its deployment context.
- Ensure appropriate human consideration and/or oversight of AI-based decisions or actions, including by establishing clear human accountability for such decisions and actions and maintaining appropriate processes for escalation and senior-leadership approval.
- Maintain appropriate processes and protections for AI operators or other personnel to report unsafe, anomalous, inappropriate, or prohibited uses to the appropriate agency channels.

- Regularly monitor and test the operation, efficacy, and risks of the AI, as well as whether current risk mitigation measures are sufficient, and make these assessments available to the AI system's operators.
- Conduct periodic human reviews at appropriate intervals to assess changes to the context, risks, benefits, and agency needs related to the AI.
- Mitigate as appropriate emerging risks from the AI that are identified through monitoring, reviews, or other mechanisms.
- Maintain appropriate processes for internal escalation and senior-leadership approval for uses of AI that pose significant degrees of risks enumerated in the AI NSM and this AI Framework, or alternatively that could harm the reputation or foreign policy interests of the United States or significantly affect international norms of behavior as determined by agency leadership.

Agencies are encouraged to apply these minimum risk management practices, and other risk management practices as relevant and applicable, to AI use cases that are not categorized as high impact, to the extent practicable and appropriate.

Additional Procedural Safeguards for AI Impacting Federal Personnel

Within 180 days of the issuance of this AI Framework, covered agencies shall ensure these practices, in addition to those practices required for high-impact AI, are followed for AI that impacts Federal personnel as described in Pillar 1(c):

- Consult and incorporate feedback from the workforce and their representatives in the design, development, and ongoing use of AI that affects them, as appropriate.
- Notify individuals and obtain consent from individuals on the use of AI that affects them, as appropriate.
- Notify individuals when the AI was used to inform an adverse employment-related decision or action that specifically concerns them, such as medical diagnosis, eligibility for employment, or access to classified information, as consistent with applicable law, regulation, and policy.
- Provide timely human consideration and potential remedy to the use of AI through a fallback and escalation system in the event that an impacted individual would like to appeal or dispute the decision informed by AI, where practicable and consistent with applicable law.

Waivers

Chief AI Officers, in consultation with the respective agency's civil liberties and privacy officers or other relevant agency officials, may waive – for a period of time not to exceed one year, but that may be renewed – one or more of the minimum-risk management practices for a specific AI application or component after making a written determination, based on a system-specific and context-specific risk assessment, that fulfilling the minimum practice would increase risks to privacy, civil liberties, or safety, or would create an unacceptable impediment to critical agency operations or exceptionally grave damage to national security.

★ ★ ★ ★ ★ ★

Chief AI Officers:

- May revoke a previously issued waiver at any time.
- Must centrally track such waivers.
- May not delegate authority to grant waivers.
- Must report to agency leadership immediately, and to the Department Head and APNSA within three days, upon granting or revoking any waiver, detailing the scope, justification, and supporting evidence.[7]
- Must reassess each waiver if there is a significant change to the conditions or context in which the AI is used.
- Must annually review all waivers granted and determine whether to reauthorize each waiver.

Civil liberties and privacy officers, in coordination with relevant agency officials, must include all waivers, detailing the scope, justification, and supporting evidence, in their annual reporting per Public Law No. 110-53 consistent with appropriate protection of sources and methods.

The Secretary of Defense on behalf of the Department of Defense, the Director of National Intelligence on behalf of the Intelligence Community, and other Department Heads, as appropriate, shall publish annually an unclassified report of the total number of waivers in their respective elements or components and how many are currently active.

Pillar III: Cataloguing and Monitoring AI Use

Inventory

Covered agencies shall conduct an annual inventory of their high-impact AI use cases, including those operating under waivers, and shall report the inventory to the APNSA. The inventory shall include at least a description of each included AI use case, its purpose and intended benefits, and the risks that such use poses and how the agency is managing those risks. Department Heads will periodically issue and update detailed instructions on the scope, timing, mechanism, and contents of the inventory.

Data Management

Within 270 days of the issuance of this AI Framework, Department Heads shall establish or update existing data management policies and procedures – including reviewing data retention policies and procedures – prioritizing enterprise applications and accounting for the unique

[7] The Secretary of Defense is permitted to produce and submit a single report to the APNSA consolidating the reports of Department of Defense components, and the Director of National Intelligence is permitted to produce and submit a single report to the APNSA consolidating the reports of Intelligence Community (IC) elements.

attributes of AI systems, with special consideration for high-impact uses of AI systems identified in this AI Framework. Department Heads shall continuously evaluate data management policies and procedures to ensure they enable responsible AI use. These updates should address at a minimum, but not be limited to, the following topics:

- Evaluating AI training data for robustness, representativeness, and reasonably foreseeable instances of harmful bias in the context of expected AI use cases.
- Recommending best practices and standardization for training data, prompts, and reviewing the quality and reliability of data post deployment.
- Handling – including documentation, management, and retention of – AI models that may have utility beyond the proximate basis for collection or are trained on information subsequently found to be inaccurate, improperly obtained, or too sensitive.
- Guidelines for using AI to make automated determinations related to mission-critical decisions.
- Guidelines for using AI in such a way that protects civil liberties, privacy, and human rights, including with regard to collection and retention of data used in AI training, as appropriate.
- Standards for AI evaluations and auditing.
- Relevant directives from the National Manager for NSS to mitigate cybersecurity risks.

Oversight and Transparency

Within 60 days of the issuance of this AI Framework, covered agencies shall appoint a Chief AI Officer if one does not already exist. The Chief AI Officer shall have the necessary skills, knowledge, training, expertise, and authority to perform the following actions with regard to AI, in addition to the responsibilities already established in section 6702(b) of the Fiscal Year 2023 National Defense Authorization Act; section 10.1(b) of E.O. 14110; section 8(c) of E.O. 13960; section 3(b) of E.O. 14091; and section 3(b) of OMB Memorandum M-24-10, as practicable:

- Serve as the senior advisor for AI to the head of the agency and other senior agency leadership and within their agency's senior decision-making forums.
- Institute the requisite governance and oversight processes to achieve compliance with the AI NSM and this AI Framework, and enable responsible use of AI in the agency, in coordination with relevant agency officials.
- Maintain awareness of agency AI activities, including through the creation and maintenance of the annual high-impact AI use case inventory.
- Work with relevant agency officials on the resourcing requirements necessary to implement the AI NSM and this AI Framework and provide recommendations on priority investment areas to build upon existing enterprise capacity.
- Advise relevant officials on improving workforce capacity and securing and maintaining the skillsets necessary for using AI to further the agency's mission and adequately manage its risks.

- Share relevant information with agency officials involved in the agency's major AI policymaking initiatives.
- Support agency involvement with appropriate interagency coordination bodies related to their agency's AI activities.
- Support and coordinate their agency's involvement in AI standards-setting bodies, as appropriate.
- Promote equity and inclusion within the agency's AI governance structures and incorporate diverse perspectives into decision-making processes.
- Work with their agency to identify and prioritize appropriate uses of AI that will advance both their agency's mission and equitable outcomes.
- Identify and remove barriers, as appropriate, to the responsible use of AI in the agency, including through the advancement of AI-enabling enterprise infrastructure, data access and governance, workforce development measures, policy, and other resources for AI innovation.
- Advocate within their agency and to the public, as appropriate, on the opportunities and benefits of AI to the agency's mission.
- Work with relevant senior agency officials to establish or update processes to measure, monitor, and evaluate the ongoing performance and effectiveness of the agency's AI applications and whether the AI is advancing the agency's mission and meeting performance objectives.
- Oversee agency compliance with requirements to manage risks from the agency's use of AI, including those established in the AI NSM and this AI Framework, and in relevant law and policy.
- Conduct risk assessments, as necessary, of the agency's AI applications to ensure compliance with the AI NSM and this AI Framework.
- Work with relevant agency officials to develop supplementary AI risk management guidance particular to the agency's mission, in coordination with officials responsible for civil liberties, privacy, and safety.
- In partnership with relevant agency officials, establish controls to ensure that their agency only uses AI that is in compliance with the AI NSM and this AI Framework.
- Provide guidance on prioritizing appropriate and responsible acquisition, deployment, use, decommissioning, and governance of AI to advance the agency's mission, including guidance related to budget and resources for AI, in coordination with relevant agency officials (such as privacy and civil liberties, authorizing, acquisition, legal, data governance, human capital, and oversight officials).

Within 60 days of the issuance of this AI Framework, covered agencies shall establish an AI Governance Board if one does not already exist, composed of relevant senior officials to govern the agency's use of AI, including assessing and, when appropriate, removing or mitigating barriers to the development and use of AI and managing its associated risks. Agencies are permitted to rely on existing relevant agency-specific policy and requirements for governance or

review bodies to fulfill this requirement, including those that are directed by E.O. 14110 and OMB Memorandum M-24-10.

AI Governance Boards must:

- Be chaired by the Chief AI Officer or an appropriate official designated by the Chief AI Officer. The full Board must convene as frequently as necessary to evaluate, on a continual basis, that AI is performing as intended.
- Include appropriate representation from senior agency officials responsible for enabling AI adoption and risk management, including at least those for information technology, cybersecurity, data, privacy and civil liberties, acquisition, budget, legal, and officials representing the agency's core mission function to which the AI will contribute.

In each covered agency, the head of the agency shall designate appropriate officials to provide oversight of the agency's AI activities within the scope of authorized duties of such officials, reporting directly to the head of the agency or the principal deputy. Roles with existing oversight functions will conduct those functions according to existing statutory authorities, such as the Privacy and Civil Liberties Officers providing oversight responsibilities as required by 42 U.S.C. Section 2000ee-1 – Privacy and civil liberties officers. At a minimum, covered agencies must have officials responsible for oversight of privacy, civil liberties, transparency, safety, and other issues related to agency AI use. The following responsibilities and tasks should be considered for officials assigned to conduct such oversight:

- Advise agency leadership, Chief AI Officers, and other relevant officials on managing risks posed by AI systems used by the agency to privacy, civil liberties, transparency, safety, and other issues determined by agency leadership.
- Develop standardized documentation for AI activities and compliance, including documentation for oversight purposes.
- Receive, review, and assess incidents of misuse identified through monitoring and evaluation of AI systems.
- Seek and consider feedback from relevant stakeholders, including civil society, technologists, academics, the private sector, and impacted communities, as appropriate.
- Report misuse identified through monitoring and evaluation of AI use – performed by the privacy and civil liberties officer or equivalent official and leveraging existing relevant processes where practicable – to their respective agency leadership, the Chief AI Officer, the Secretary of Defense, and the Director of National Intelligence, as applicable, and to the APNSA, as appropriate.
- Ensure that the agency has adequate procedures to receive, investigate, respond to, and redress complaints about the agency's use of AI, to include procedures to receive complaints anonymously, when appropriate.
- Perform responsibilities confidentially, when appropriate, so that agency personnel may raise concerns without fear of reprisal; confidentiality, however, shall not extend to significant misconduct, to include violations of law or government ethics, or when otherwise precluded by law.
- Document AI misuse, AI incidents, and lessons learned.

Transparency is essential to earning and retaining public trust. Consistent with this goal, the privacy and civil liberties officer or other relevant oversight official shall periodically, but not less frequently than annually, submit a report on their activities associated with AI oversight, including, for example, evaluations of risk management processes. The report shall be delivered to, at a minimum, the head of the agency. It shall:

- Be in unclassified form to the greatest extent practicable, with a classified annex when necessary.
- Integrate into existing reporting requirements on privacy and civil liberties, such as Section 803 of the Implementing Recommendations of the 9/11 Commission Act of 2007 (Public Law No. 110-53), as applicable, for issues related to privacy and civil liberties.
- Be made available to the public to the greatest extent that is consistent with the protection of classified or controlled information, and applicable law.

Department Heads shall ensure that relevant oversight officials have sufficient information, expertise, training, and adequate funding to effectively carry out these functions.

Pillar IV: Training and Accountability

Covered agencies shall establish standardized training requirements and guidelines for the workforce on the responsible use and development of AI, including AI training for privacy and civil liberties officers, risk management training for officials responsible for deciding to deploy or develop AI systems, and relevant training for AI developers, operators, users, supervisors, and consumers of AI outputs.

Covered agencies shall update their policies and procedures, as needed, to ensure adequate accountability for those involved in the development, deployment, and use of AI. No AI systems should be deployed or used without adequate applicable policies and procedures in place, such as approvals by appropriate officials, to enable adequate accountability. Updated policies shall:

- Identify who in the AI lifecycle assesses risks associated with the use of AI, including, but not limited to, information technology-, legal-, policy-, security-, privacy-, civil liberties-, and safety-related risks.
- Establish appropriate mechanisms to hold relevant personnel, including AI developers, operators, and users, accountable for their contributions to and use of AI system decisions and action.
- Require appropriate documentation and reporting, including as directed in this AI Framework.
- Provide processes for reporting incidents of AI misuse, investigations of reported incidents, and processes for taking corrective actions.

Agencies shall update whistleblower protections as appropriate to clarify procedures for AI systems, which shall ensure that all personnel who use AI as a component of NSS or otherwise for military and intelligence purposes can report concerns about AI improperly harming civil liberties, privacy, and safety to relevant oversight officials.