CNA

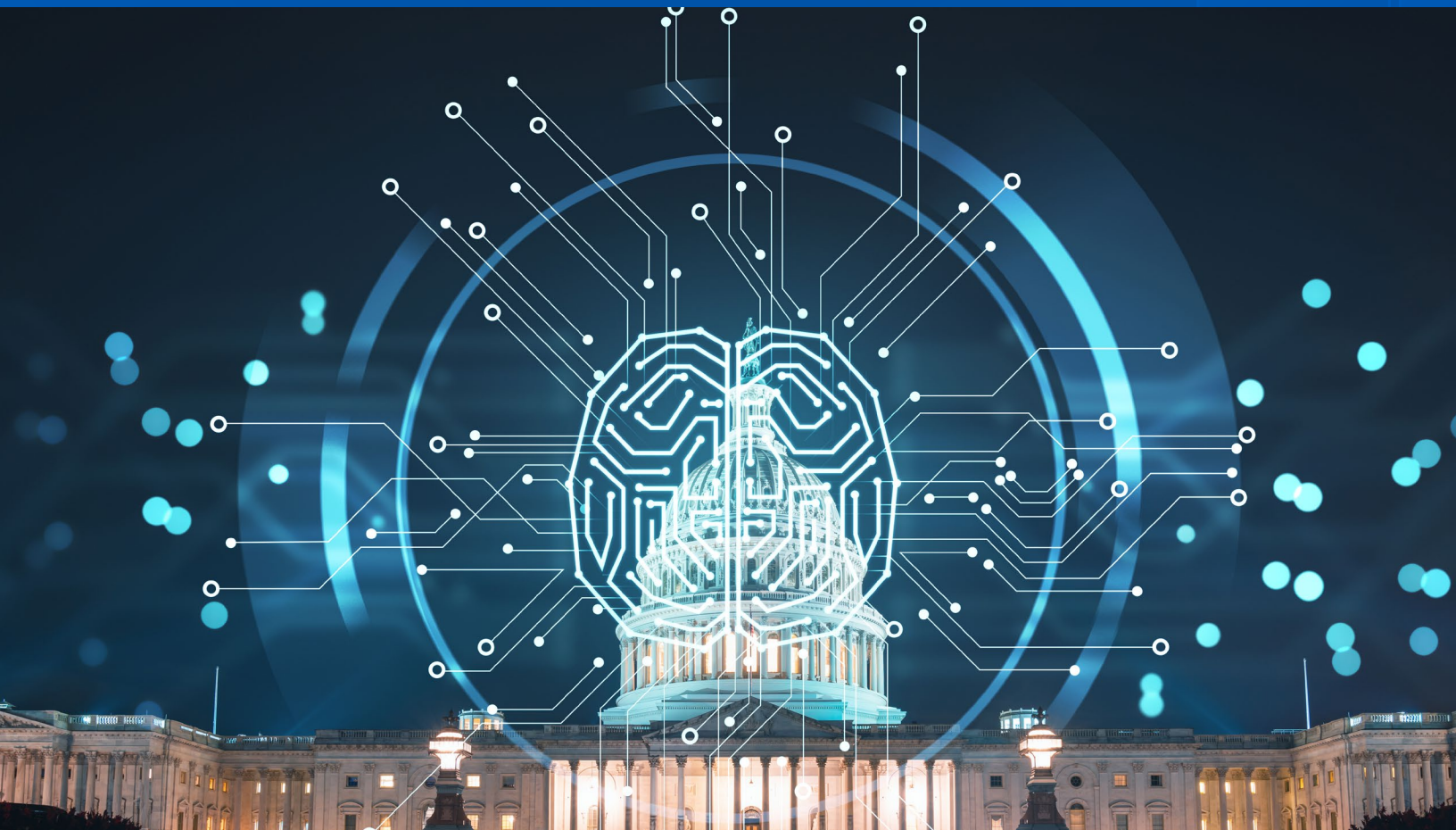# CNA's Artificial Intelligence (AI) Maturity Model for Government Agencies

James Baney, John Mahoney, Kevin Inks, and Yee San Su

Approved for public release. Unlimited distribution.

CNA

Approved by: **May 2025**

David Kaufman, Vice President
Safety and Security Division
CNA's Institute for Public Research

Introduction

Within the past few years, emerging artificial intelligence (AI) systems have reshaped a variety of tasks and offered exciting opportunities to expand capabilities across a wide set of fields. These systems are likely to be increasingly useful to government agencies in a variety of applications, including image recognition and textual analysis. Federal and state governments have begun to formulate their approaches for governing AI and using it effectively and fairly. CNA's AI maturity model offers a framework to help agencies at all levels of government evaluate the maturity of their AI capabilities, plan for the development of these capabilities, and prioritize their efforts within an increasingly complex and rapidly advancing AI field.

Although emerging AI technologies have become commonly used in private industry, academia, and private homes only recently, AI systems have long been used in specialized technical applications. Government agencies may already have capabilities in AI, broadly defined by 15 U.S.C. 9401(3) as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments." There are inevitable challenges in harmonizing emerging and already-existing capabilities, identifying the most promising use cases for these capabilities, managing risks, and organizing agency-wide structures for the efficient and safe use of AI. This maturity model for government agencies' AI capabilities serves as a blueprint from which government agencies can approach these issues, understand their current AI capabilities, and chart a course for future directions.

Organization

The maturity model is meant to be comprehensive, accounting for the full scope of government agency AI activities. It is organized into five *domains* that describe the highest level components of agencies' AI maturity:

1. Productive governance models are established: The agency's AI systems are subject to a robust governance system and human oversight and are in compliance with all relevant laws and statutes.

2. Efforts are sufficiently resourced: The agency's AI initiatives possess the necessary financial, hardware, software, data, and human resources.

3. Outputs are impactful: The agency has established performance standards, testing procedures, and policies to ensure that AI systems are applied to suitable use cases and contribute to the agency's overall mission.

4. Products and results are trustworthy: The agency designs and manages AI systems so that they are representative, transparent, and unbiased.

5. Products are safe and secure: The agency effectively manages AI-related risks and protects AI systems and the data they contain from malicious attack or inappropriate disclosure.

Domains are grouped into *subdomains*, each of which describes a more tangible and specific element of AI maturity, such as funding, testing and evaluation, or cybersecurity. These subdomains cover a broad set of issues relevant to AI maturity and were designed to be as exhaustive, actionable, and distinct as possible. Subdomains are further divided into *topics*, which are the most detailed components of agency AI capabilities in this model.

Agencies' capabilities can be assessed for each topic according to *stages of maturity*. As shown in Figure 1, the model uses four stages:

1. Developing: The agency has only limited capabilities in this area, implemented incidentally or by employees working independently and without overarching governance, structure, or strategy.

2. Performed: The agency's capabilities in this area are guided by some level of intentionality and strategy but remain underdeveloped.

3. Established: The agency's capabilities in this area are fully developed, but opportunities for continued improvement remain underexplored.

4. Optimized: The agency's capabilities in this area are fully developed and subject to consistent, systematic efforts to improve, strengthen, and expand them.

For each topic, the model provides several indicators that describe how government agencies' AI operations may be performed for each maturity level. These indicators serve as the benchmark for maturity assessments: agencies will compare their own operations to the indicators to identify their levels of maturity.

Agencies operating at high levels of maturity will have several AI plans, policies, processes, and organizational structures in place. The model refers to these products and structures that support AI operations as *tangible outputs*. The model summarizes the tangible outputs associated with high maturity levels for each subdomain.

Figure 1. Stages of maturity
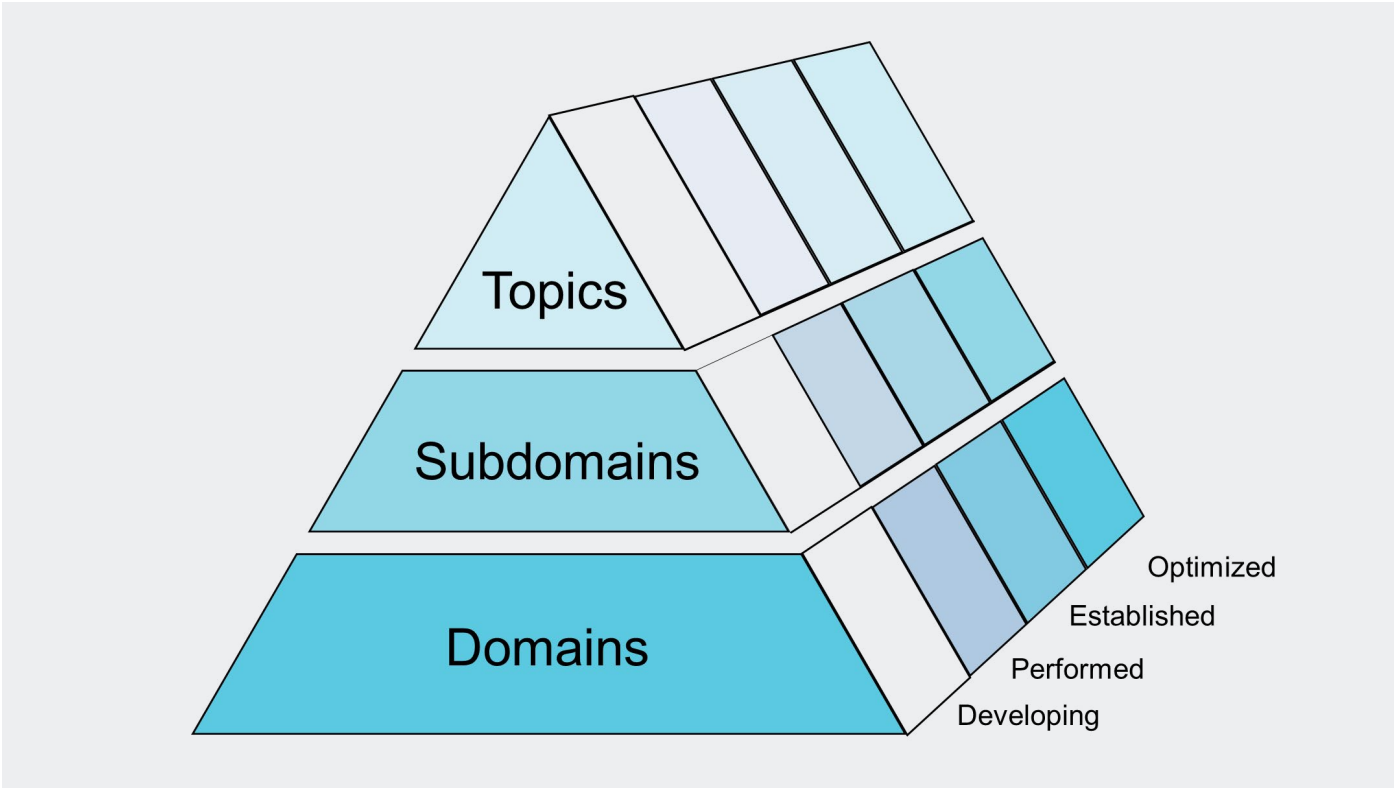
Source: CNA.

Methodology

CNA developed this maturity model based on current policies, plans, and guidance related to AI from a variety of organizations. During our analysis, we took the following steps:

- Assembled a corpus of 39 publicly available reference documents that included strategies, policies, plans, and guidance from the federal interagency, individual federal departments and agencies, state governments, international organizations, nongovernmental organizations, academia, and the private sector.

- Categorized the content of key documents and used the categories to draft tentative domains and subdomains intended to provide a comprehensive classification of government agency AI capabilities.

- Coded content from the remaining reference documents to the domains and subdomains, updating the domains and subdomains as needed and sharpening the distinctions between them as we went.

- Developed the four maturity levels used in the model along with the characteristics of each.

- Used the content from the reference documents to develop tentative maturity indicators for each subdomain and for each maturity level. In doing so, we divided the subdomains into the topics that are included in the maturity model.

- Performed internal reviews of the tentative maturity indicators, including comparing them to other AI-related maturity models from various sources to identify any areas of AI capability we had potentially missed. We left this review of these external maturity models until the end to avoid biasing our model development process. We did not identify any areas of AI capability that we had missed.

- Provided the draft maturity model to CNA subject matter experts outside of the project team for their review and updated the model accordingly.

Using the Model

This model should be used as a tool for government agencies to understand their AI capabilities as well as to inform their further development, policy decisions, and internal operations. It should be used qualitatively as a map by which the agency can place itself in the overall landscape of AI capabilities and orient itself toward further progress.
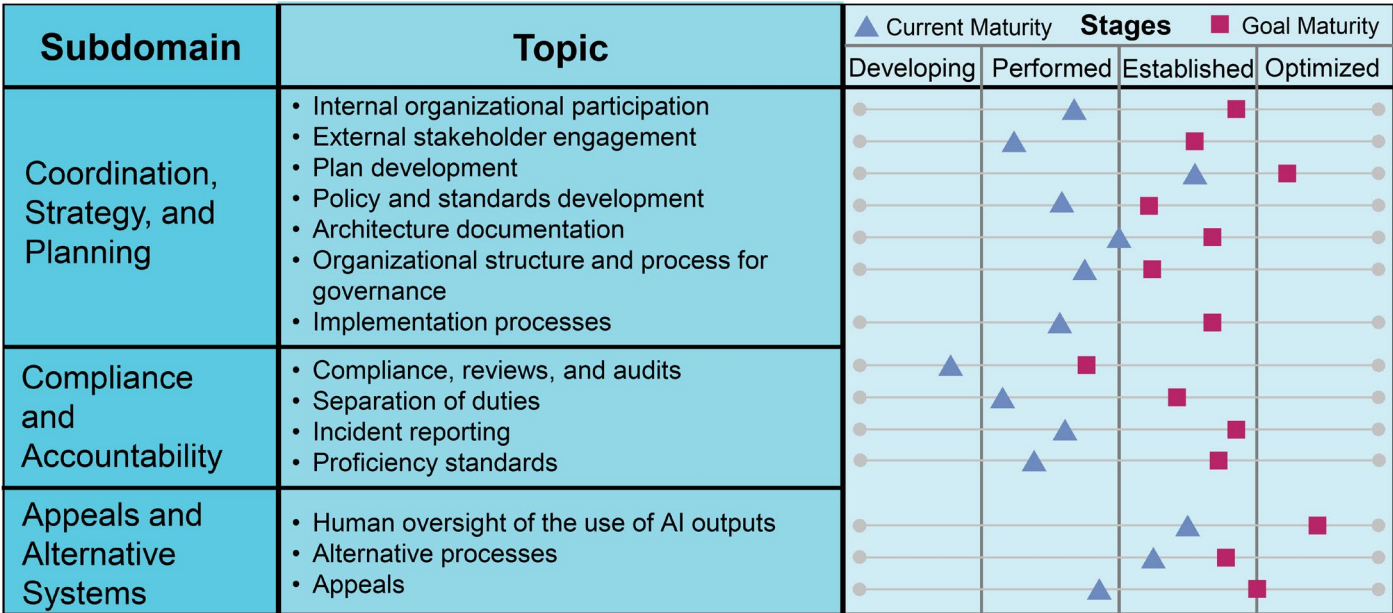
The agency should evaluate its maturity level in each domain, subdomain, and topic by analyzing which set of indicators most closely resembles its own AI operations. This diagnostic exercise will give the agency a holistic sense of its general level of maturity with respect to AI. It can use the maturity indicators of the next highest domain to focus specific capability-building activities.

Certain elements of the model may be more salient for some agencies than for others. Depending on its specific mission and goals, the agency may wish to focus more energy on developing capability in specific domains or subdomains because of their perceived importance or due to preexisting plans and ongoing activities in those areas.

Figure 2 illustrates a typical use of the maturity model. Within each domain, subdomain, and topic, agencies can self-assess their current levels of maturity (notionally indicated by the blue triangles) and their desired future maturity levels (notionally indicated by the red squares). Differences between desired future maturity levels and current maturity levels indicate capability gaps, and the maturity indicators associated with each topic offer tangible milestones associated with closing the gaps. Agencies can prioritize the gaps they intend to address and develop plans to increase their capabilities in those areas.

Figure 2. Typical use of maturity model

Subdomain	Topic	▲ Current Maturity	**Stages**	■ Goal Maturity	
		Developing	Performed	Established	Optimized
Coordination, Strategy, and Planning	• Internal organizational participation • External stakeholder engagement • Plan development • Policy and standards development • Architecture documentation • Organizational structure and process for governance • Implementation processes				
Compliance and Accountability	• Compliance, reviews, and audits • Separation of duties • Incident reporting • Proficiency standards				
Appeals and Alternative Systems	• Human oversight of the use of AI outputs • Alternative processes • Appeals				

Source: CNA.

Maturity Model

Domain 1: Productive governance models are established

Domain 1: Productive governance models are established	
The agency's AI systems are subject to a robust governance system and human oversight and are in compliance with all relevant laws and statutes.	
Subdomain	**Definition**
Coordination, strategy, and planning	The development and use of the agency's AI systems is planned strategically and executed in coordination with appropriate internal and external stakeholders.
Compliance and accountability	The agency is compliant with relevant laws and regulations and has a system of standards and policies that enable internal and external stakeholders to hold the agency accountable for its use of AI.
Appeals and alternative systems	The agency's AI systems are subject to human oversight, and their decisions can be appealed.

Subdomain 1.1: Coordination, strategy, and planning

Definition
The development and use of the agency's AI systems is planned strategically and executed in coordination with appropriate internal and external stakeholders.

Topics and goals
1. **Participation of individuals inside the organization.** The agency draws on a variety of individuals—both those with technical expertise and those with nontechnical expertise—to support its AI systems across their life cycles.

2. **Engagement with external stakeholders.** The agency effectively incorporates perspectives from communities and organizations affected by the AI systems it employs and interacts with interagency, community, industry, and other partners to refine its use of AI.

3. **Plan development.** The agency can articulate how it will approach using AI systems as well as what it intends to achieve through its use of AI.

4. **Policy and standards development.** The agency has identified AI-related roles and responsibilities, set standards for AI development and ethics, and determined how and when its staff may use AI systems.

5. **Architecture documentation.** The agency has documented its usage of AI systems, including the models it uses, the tasks those models perform, and the current and desired states of those systems.

6. **Organizational structure and process for governance.** The agency has established an AI-specific governance model and identified key AI governance–related roles and responsibilities and internal AI-related organizations, such as an AI governance board and an AI community of practice.

7. **Implementation process.** The agency can clearly characterize the readiness of its AI systems, describe the assumptions underlying those systems, identify the objectives of its AI systems, and specify the conditions under which an AI system will be retired.

Tangible outputs

Policies, plans, and guidance

- Guidance for the agency's partners (including at other levels of government) and other stakeholders on AI-related issues within the agency's mission area

- Policies that identify AI-related roles and responsibilities across the agency

- Guidance on how agency staff can (or cannot) use generative AI tools

- Standards for topics related to AI development, deployment, use, and ethics

- A framework for how to describe the readiness of the agency's AI systems

- Policies that specify where and for how long retired AI systems, models, data, and associated artifacts will be stored

People and organizations

- A chief AI officer

- An AI governance board

- An agency-wide AI community of practice

Products and processes

- An AI systems inventory

- A strategy documenting the agency's approach to using AI systems

- An AI governance model

Subdomain 1.1: Coordination, strategy, and planning

The development and use of the agency's AI systems is planned strategically and executed in coordination with appropriate internal and external stakeholders.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
1.1.1 Participation of individuals inside the organization	☐ AI system planning, development, implementation, and monitoring is performed only by the agency's technical staff (e.g., data scientists and software developers).	☐ Nontechnical stakeholders (such as stakeholders from the agency's privacy, legal, and risk management offices, along with system operators and associated union officials) sometimes participate in AI system planning, development, implementation, and monitoring, but they do so sporadically and on a project-by-project basis.	☐ The agency has documented expectations that both technical and nontechnical stakeholders will participate in AI systems efforts throughout the system life cycle, and this is consistently the case across the agency's AI projects.	☐ The agency can identify numerous examples in which input from nontechnical stakeholders has had significant effects on plans for AI systems.
1.1.2 Engagement with external stakeholders	☐ The agency has not participated in forums where other, similarly situated agencies share their approaches to AI-related challenges and opportunities. ☐ The agency has not engaged with industry partners or academia on applications of AI-to-agency functions.	☐ The agency has participated in some interagency AI forums, although the objectives of these engagements have been unclear. ☐ Engagements with members of industry and academia on AI topics occur sporadically and are driven by individuals or specific offices or projects.	☐ The agency regularly participates in interagency forums with other, similarly situated agencies, and has prioritized specific topics for coordination and information sharing. ☐ The agency regularly engages with industry or academia partners who are developing AI tools in a coordinated way.	☐ The agency can identify specific areas in which information learned via participation in interagency forums has had significant effects on its own AI activities. ☐ Agency engagements with members of industry and academia have contributed to innovations in applying AI to agency activities and have brought identifiable benefits to the agency.

1.1.2 Engagement with external stakeholders	☐ The agency has not attempted to identify its role in providing guidance to partners (including at other levels of government) and other stakeholders on the use of AI.	☐ The agency has determined what guidance it should provide to its partners and other stakeholders on AI matters, although it has not yet acted to produce the guidance.	☐ Where appropriate, the agency has developed guidance on AI matters for its partners and other stakeholders.	☐ The agency has collected feedback from partners and stakeholders on guidance it has provided them and has used that feedback to improve its offerings.
1.1.3 Plan development	☐ The agency has not developed an agency-wide AI strategy. ☐ The agency has not integrated AI into its key IT plans and procedures, such as its cybersecurity and risk management strategies. ☐ The agency has not defined its approach to centralizing AI systems, infrastructure, and resources versus federating them across the agency. By default, individual offices and programs address their own AI needs, without coordination across the agency.	☐ The agency has published a high-level AI strategy but has not addressed key issues for strategy implementation, such as when and how the agency will use AI and how it will ensure responsible AI use. The strategy is not accompanied by an associated implementation plan. ☐ The agency has made some efforts to integrate its AI strategy with its other key IT plans and procedures, but these efforts have been unfinished or insufficient or are otherwise incomplete. ☐ The agency has documented some individual decisions about whether to centralize or federate some AI efforts, but the overarching strategy is unclear to most stakeholders.	☐ The agency has an AI strategy with clear goals, objectives, and measurement approaches. The strategy and associated plans define a clear approach to the agency's use of AI, including when and how the agency will use AI, what the high-impact use cases are, how AI will support the agency's mission and critical functions, and how the agency will ensure it uses AI responsibly. ☐ The agency's AI strategy is integrated with its other strategies, including its cybersecurity and risk management strategies. ☐ The agency has a documented approach that describes how it will balance centralizing AI and using shared services versus federating AI implementation throughout the agency.	☐ The agency regularly reviews and updates its AI strategy and associated implementation plans, and it reports on its progress relative to them. The agency updates plans and activities based on these reviews. ☐ As the agency reviews and updates its IT and AI strategies, it continues to ensure that they are well integrated with each other. ☐ The agency regularly reviews proposed long-term AI investments and planning for consistency with its overall AI strategy.

1.1.4 Policy and standards development	☐ The agency has not documented the roles and responsibilities of its staff with respect to AI. ☐ The agency has not identified AI standards or developed them. ☐ The agency has not issued policy or guidance on how staff may and may not use generative AI tools in their work. ☐ The agency does not assess compliance with AI policies and standards, or does so very infrequently.	☐ The agency has developed policy that specifies AI-related roles and responsibilities, but significant numbers of staff in positions of responsibility are poorly prepared to carry out their assigned responsibilities. ☐ The agency has prioritized an initial set of AI standards that it needs to develop and has taken initial steps to develop them. ☐ The agency has recently provided policy or guidance on how staff may and may not use generative AI tools, but compliance with the policy is inconsistent or unknown. ☐ The agency can provide quantitative information about compliance with some policies.	☐ The agency has enacted policies and issued associated guidance that identify the roles and responsibilities of its staff with respect to AI. Agency staff are broadly familiar with the policies and guidance and feel well positioned to carry out their responsibilities under them. ☐ The agency has developed and approved standards that cover topics within AI development, deployment, use, and ethics. ☐ The agency has issued clear policy or guidance on how staff may and may not use generative AI tools and monitors compliance with such policy or guidance. ☐ The agency consistently tracks compliance with its AI policies. Compliance is consistently high throughout the agency.	☐ The agency periodically reviews its policies and guidance that define AI-related roles and responsibilities based on staff experience and updates them where appropriate. ☐ The agency undertakes regular reviews of its AI-related standards and their implementation and can identify deliberate actions that governing bodies have undertaken in response to identified gaps in standards. ☐ The agency has a process for reviewing and updating policies and guidance on staff use of generative AI as new use cases and technologies emerge. ☐ The agency tracks policy exceptions and escalations and uses the data to evaluate policy needs and make policy updates.
1.1.5 Architecture documentation	☐ The agency has not documented its AI assets, including its tools, infrastructure, and processes, or how they are used.	☐ The agency has documented the current state of its AI tools, infrastructure, and processes but has not documented their desired future states.	☐ The agency has documented both the as-is and desired states of its AI tools, infrastructure, and operations.	☐ The agency regularly uses its enterprise architecture documentation to assess its proposed AI investments.

1.1.5 Architecture documentation	☐ The agency does not have an AI systems inventory program and is unable to comprehensively identify the AI systems it employs.	☐ The agency has a centralized AI systems inventory, but the information it collects about its AI systems is minimal and the inventoried systems are incomplete.	☐ The agency has a comprehensive AI systems inventory that allows for rapid identification of key information about agency AI systems, including the types of models they use, the tasks they perform, their degree of autonomy, the points of contact for them, and the breadth of their deployment.	☐ The agency regularly reviews the completeness, usability, and efficacy of its AI systems inventory and takes specific actions to address gaps.
1.1.6 Organizational structure and process for governance	☐ The agency has not implemented a governance model specific to making decisions about AI development and use. AI investment and implementation decisions are made by individual offices and programs. ☐ The agency has not designated a chief AI officer.	☐ The agency has implemented an AI governance model, but it can identify very few instances in which governance processes have positively affected AI outcomes for the agency. ☐ The agency has designated a chief AI officer, although that officer is still in the process of establishing the value to be added from their role.	☐ The agency's AI governance model has clearly defined goals and objectives, and the agency can identify numerous examples of how its governance processes have contributed to achieving them. ☐ The agency's chief AI officer has demonstrated clear value through activities such as leading the development of enterprise AI strategy and providing AI expertise, promoting AI innovation across the agency, overseeing compliance with AI requirements, implementing an enterprise AI risk management approach, and facilitating the procurement of AI tools and systems.	☐ The agency has assessed the performance of its AI governance model and made updates to address deficiencies as needed. ☐ The agency's AI community of practice or similar forums are broadly recognized as a key driver of the agency's AI advances.

1.1.6 Organizational structure and process for governance	☐ The agency has not established an AI governance board composed of senior leaders. ☐ The agency does not have an AI community of practice, and there is little information sharing between AI practitioners in different offices across the agency.	☐ The agency has an AI governance body composed of senior leaders, although the body's effect on the agency's AI operations has been minimal. ☐ The agency has established an AI community of practice or similar forum for exchanging information between AI practitioners, although few participants can identify specific instances in which they have obtained useful information through it.	☐ The agency can identify several ways in which its AI governance board has contributed to improved outcomes in its use of AI. ☐ Participants in the agency's AI community of practice widely recognize the value of their participation and can identify specific cases in which they have used information learned in it.	
1.1.7 Implementation process	☐ The agency has not established a consistent framework to characterize the readiness of its AI systems (e.g., research, proof-of-concept, pilot, and production implementation stages). ☐ The agency rarely or never documents the operational conditions, assumptions, and limitations associated with its AI systems. ☐ The agency's AI systems rarely have clear, measurable goals and objectives.	☐ The agency has established a consistent framework to characterize the readiness of its AI systems, but use of that framework is inconsistent. ☐ Agency policy requires documentation of the operational conditions, assumptions, and limitations associated with its AI systems, although compliance with the requirement is inconsistent. ☐ The agency requires documentation of the goals and objectives of its AI systems but typically does not evaluate whether the systems achieve their goals.	☐ The agency uses a consistent framework to describe the readiness of its AI systems and can characterize where its AI systems are on that path during their development and deployment. ☐ The agency consistently documents the operational conditions, assumptions, and limitations associated with its AI systems. ☐ The agency consistently identifies clear, measurable goals and objectives of each AI system and consistently evaluates system performance against objectives.	☐ The agency can identify instances during which it has used a technological readiness roadmap to inform AI investment decisions. ☐ The agency can identify several cases in which reviews of system performance against objectives have led to specific actions for system improvements.

| 1.1.7 Implementation process | ☐ Decisions on retiring AI systems are typically made by individual programs or organizations without predefined criteria. | ☐ The agency has some requirements for defining conditions for retiring AI systems, although compliance with those requirements is inconsistent. | ☐ The agency consistently identifies the conditions under which it will retire AI systems as part of the initial system planning process and retires systems when those conditions are met. Policies specify where and how long retired AI systems, models, data, and associated artifacts are stored. | |

Subdomain 1.2: Compliance and accountability

Definition
The agency is compliant with relevant laws and regulations and has a system of standards and policies that enable internal and external stakeholders to hold the agency accountable for its use of AI.

Topics and goals
1. **Compliance, reviews, and audits.** The agency can ensure that its AI systems are compliant with relevant laws, regulations, policies, and standards.

2. **Separation of duties.** The agency has divided AI-related responsibilities among different roles to ensure AI accountability.

3. **Incident reporting.** The agency is able to receive reports from internal and external stakeholders about AI-related incidents and concerns and can route these issues to individuals with the relevant expertise.

4. **Proficiency standards.** The agency can assess and ensure the proficiency of its AI system operators.

Tangible outputs
Policies, plans, and guidance
- Guidelines for performing AI-related compliance reviews

Products and processes
- Controls that help the agency ensure that its AI systems meet relevant compliance requirements

- A strategy describing how and when the agency will use third-party audits to assess the compliance of its AI systems

Subdomain 1.2: Compliance and accountability

The agency is compliant with relevant laws and regulations and has a system of standards and policies that enable internal and external stakeholders to hold the agency accountable for its use of AI.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
1.2.1 Compliance, reviews, and audits	☐ The agency has not identified the laws, regulations, policies, and standards that its AI systems must comply with. Individual offices have varying degrees of awareness of such requirements, and their understanding of them is frequently incomplete. ☐ The agency has reviewed very few of its AI systems for compliance with relevant laws, regulations, policies, and standards, and the degree of compliance across different systems is not well understood. ☐ The agency has not developed an approach for how or when to use third-party audits and assessments to ensure that its AI systems are in compliance with laws, regulations, policies, and standards. As a result, it has not conducted third-party audits or assessments of its systems.	☐ Requirements for legal, regulatory, policy, and standard-based requirements have been compiled, but awareness and understanding of them among appropriate staff is uneven. ☐ The agency has performed compliance reviews for some systems, but it has not defined a standard framework or structure for such reviews. ☐ Although the agency can demonstrate compliance with laws, regulations, policies, and standards for several of its systems, a substantial fraction of systems are known to be out of compliance.	☐ Legal, regulatory, policy, and standards-based requirements involving AI are understood, managed, and documented, including requirements specific to the agency's mission area. ☐ The agency has centrally managed controls in place to ensure its AI systems meet requirements (such as identifying the documentation required to demonstrate compliance). These controls allow the agency to understand the compliance status of almost all of its AI systems. ☐ The agency has established guidelines for which systems must be reviewed, the frequency of AI system reviews, the level of detail to be reviewed, and the conditions under which systems must be re-reviewed.	☐ The agency has an established process for ensuring the understanding and documentation of law, regulations, policies, and standards when new requirements or changes to requirements are introduced. Agency staff are formally trained and retrained on requirements as appropriate. ☐ The agency can demonstrate a history of rapidly addressing deficiencies associated with legal, regulatory, policy, and standards requirements.

1.2.1 Compliance, reviews, and audits		☐ Some agency AI systems have undergone third-party audits and assessments, although such audits and assessments are performed on a program-by-program basis and are not part of a documented strategy for ensuring compliance.	☐ The vast majority of the agency's systems have been demonstrated to be compliant with laws, regulations, policies, and standards. ☐ The agency has documented a strategy for how and when it uses third-party audits and assessments to enhance compliance, and it has had numerous AI systems undergo such audits and assessments.	
1.2.2 Separation of duties	☐ There is little or no "separation of duties" across AI development, operation, and oversight roles in the agency. For example, the same individuals are frequently responsible for the development and testing of systems, or for the operation and monitoring of systems.	☐ The agency has taken steps to separate duties for AI development, operation, and oversight, but efforts are nascent or unenforced.	☐ The agency's policies require a separation of duties among development, operation, and oversight roles, and in practice such roles are consistently separated across the agency.	☐ The agency periodically reviews how different AI roles and responsibilities are performed to ensure that there is an adequate and efficient separation of AI-related duties.
1.2.3 Incident reporting	☐ The agency does not have a standard process for reporting AI incidents or a framework for managing incidents.	☐ The agency has a formal process for reporting AI incidents, but in practice the process has been followed only infrequently.	☐ The agency has a centrally managed framework for reporting and managing AI incidents or controversies, focusing on those that are negative or harmful. The incident reporting and management process is regularly used when appropriate.	☐ The agency regularly reviews trends in AI incidents and can identify specific activities it has undertaken to prevent future incidents as a result of these reviews.

1.2.3 Incident reporting	☐ The agency does not have a "whistleblower" process to allow for the protected reporting of AI system concerns.	☐ The agency has formal whistleblower processes and resources, but most staff and stakeholders are unfamiliar with them.	☐ The agency has developed whistleblower policies and processes that facilitate reporting of serious AI system concerns and regularly publicizes the policies to appropriate stakeholders. Processes include alerting responsible AI staff when such incidents are reported or when their expertise may be necessary.	
1.2.4 Proficiency standards	☐ The agency does not employ proficiency standards for AI system operators.	☐ Some of the agency's AI systems use proficiency standards for operators.	☐ Where appropriate, processes and associated certifications for operator and practitioner proficiency are defined and documented for most of the agency's AI systems.	☐ The agency periodically reassesses its standards for AI system operations and can identify updates made as a result of such assessments.

Subdomain 1.3: Appeals and alternative systems

Definition

The agency's AI systems are subject to human oversight, and their decisions can be appealed.

Topics and goals

1. **Human oversight of the use of AI outputs.** The agency has established when, and in what ways, human operators can overrule AI system outputs.

2. **Alternative processes.** The agency provides users with alternative, non-AI systems that can be used when necessary.

3. **Appeals.** The agency has an established process through which AI system users can appeal AI decisions.

Tangible outputs

Policies, plans, and guidance

- Guidance on determining the authorities human operators have in overruling AI system outputs

- Rules that specify the appropriate level of authority human operators have in overruling AI system outputs

Products and processes

- Alternative processes to the agency's AI systems

- A process through which users can appeal decisions made by the agency's AI systems

Subdomain 1.3: Appeals and alternative systems

The agency's AI systems are subject to human oversight, and their decisions can be appealed.

	Stage of Maturity			
Topic	Developing	Performed	Established	Optimized
1.3.1 Human oversight of the use of AI outputs	☐ Expectations for what type of human oversight is required for different AI systems, as well as what authorities humans have when reviewing AI outputs, are unclear for the vast majority of the agency's AI systems.	☐ Although authorities and expectations for how human operators interpret and potentially overrule AI outputs have been established for some systems, their use is inconsistent across the agency.	☐ Authorities and expectations for humans with respect to interpreting and overruling AI outputs are consistently well established across AI systems. ☐ The agency uses an established risk framework to determine appropriate levels of authority for humans to interpret and overrule AI system outputs.	☐ The agency collects data on instances in which humans overrule AI system results and uses the data to evaluate AI performance and inform AI improvements.
1.3.2 Alternative processes	☐ The degree of choice that stakeholders have in whether they are subject to AI systems is not clear across the agency. ☐ Mechanisms for opting out of processes that use AI systems are generally not established. ☐ AI system performance monitoring that can trigger alternative processes for high-risk cases is not performed.	☐ The agency is inconsistent across its AI systems in defining the degree of choice that stakeholders have in being subject to AI systems. ☐ Opt-out processes are defined for some AI systems, but definitions are inconsistent across the agency. ☐ The agency has implemented automatic triggers of alternative processes for some AI systems but lacks central policy or guidance for when this is appropriate.	☐ The agency consistently identifies the degree of choice that stakeholders have in whether they are subject to AI systems. ☐ Alternative processes to the use of AI systems are consistently defined. ☐ AI system monitoring triggers the implementation of alternative processes when risk metrics exceed tolerances.	☐ The agency collects data on instances in which alternative processes are used, compares AI results with the results of alternative processes, and uses results to evaluate AI performance and inform AI improvements.

1.3.3 Appeals	☐ The agency does not have mechanisms in place for those negatively affected by AI determinations to appeal the results. AI determinations are irreversible. ☐ Cases in which appeals are allowable are largely undefined. ☐ Appeals processes are not clearly communicated to those affected by AI determinations.	☐ Some of the agency's AI systems allow for the appeal and reversal of AI-generated results, but this is inconsistent across the agency. ☐ Cases in which appeals are allowable are defined for some of the agency's AI-based processes. ☐ Communications of appeals processes are left to individual programs and are implemented only for some of the agency's systems.	☐ Appeals processes are well established and consistently include remedy and redress procedures for those affected by problematic AI outputs. ☐ Cases in which affected persons can appeal AI results are consistently well defined. ☐ Mechanisms for appeals are consistently made clear to those affected by AI determinations.	☐ The agency collects data on appeals and has used those data to inform AI system improvements.

Domain 2: Efforts are sufficiently resourced

Domain 2: Efforts are sufficiently resourced	
The agency's AI initiatives possess the necessary financial, hardware, software, data, and human resources.	
Subdomain	**Definition**
Datasets	The agency possesses and makes readily available data of sufficient quality for use in developing and testing systems.
Infrastructure	The agency possesses the necessary software and hardware infrastructure to develop and use AI systems.
Procurement	The procurement of AI systems is systematic.
Workforce and expertise	The agency is able to recruit, retain, and develop personnel with AI expertise as well as foster an AI-literate workforce.

Subdomain 2.1: Datasets

Definition
The agency possesses and makes readily available data of sufficient quality for use in developing and testing systems.

Topics and goals
1. **Data quality.** The agency is able to ensure that training and testing data are complete, accurate, valid, and appropriate for the AI systems that use them.

2. **Data accessibility.** The agency's datasets are usable and accessible to authorized individuals.

Tangible outputs

Policies, plans, and guidance
- Guidance on ensuring the quality of training and testing data

- Standards for data exchange and use

Products and processes
- Procedures for identifying and correcting erroneous data

- An enterprise data management program

- A central repository for AI-related datasets

- Processes to manage access to sensitive or restricted data

Subdomain 2.1: Datasets

The agency possesses and makes readily available data of sufficient quality for use in developing and testing systems.

	Stage of Maturity			
Topic	Developing	Performed	Established	Optimized
2.1.1 Data quality	☐ The agency does not have procedures in place to ensure that the data used as inputs to, or training material for, AI models are complete, accurate, valid, or appropriate. The quality of data used by AI systems is often poorly understood. ☐ The agency does not have processes in place to identify or correct problematic data. ☐ The agency has not identified data stewards or similar positions responsible for ensuring the quality of datasets used in AI systems.	☐ Some offices or programs that manage AI systems attempt to ensure that input and training data are complete, accurate, valid, and appropriate for their intended use cases, but there are no agency-wide procedures or policies to guide those efforts. ☐ The agency may attempt to identify and correct problematic data, but it does not do so in a systematic or policy-guided way. ☐ There are some individuals or offices that function as de facto data stewards, but their roles are not formally assigned or defined by the agency.	☐ The agency has procedures, standards, or guidance in place to ensure that input and training data are complete, accurate, valid, and appropriate for the intended use cases of AI systems. The agency has performed quantitative evaluations of the data used by its AI systems and can demonstrate that the data used in most of its AI systems are appropriate. ☐ Procedures are in place to identify and correct problematic data. ☐ The agency has an enterprise data management program that employs modern data management practices and includes data stewards or other individuals responsible for the quality of its datasets. The agency can identify improvements to data quality that have resulted from its data management program.	☐ The agency tracks trends in observed data quality issues and establishes procedures, standards, and guidance to address recurring challenges. ☐ The agency has assessed the effectiveness of its data management program with respect to AI applications and tailored program activities accordingly.

2.1.2 Data accessibility				
	☐ The agency does not possess a central repository for datasets at any organizational level, and identifying available data for AI applications is a common challenge across the agency. ☐ Nonstandard data structures frequently hinder the exchange and use of data or require extensive data wrangling by AI analysts. ☐ There are no systematic, organization-wide processes in place to review and facilitate user access to restricted data. As a result, the agency has experienced challenges with improper access to restricted data and/or excessive restrictions on data access.	☐ The agency does not possess a single, agency-wide, central repository for datasets, but there may be semicentralized datasets for subcomponents of the agency or informal interoffice communication so that individuals are frequently aware of where datasets are located. ☐ Some standards for data exchange facilitate the exchange and use of data, but their use is uneven across the agency. ☐ Some offices and programs have processes in place to review and facilitate user access to restricted data, but they are not harmonized across the institution.	☐ Most of the agency's datasets are discoverable and accessible (for example, through inclusion in a central repository) and are accompanied by explanatory metadata. ☐ Standards that facilitate data exchange and use are commonly employed across the agency. ☐ Common processes are in place to review and facilitate user access to restricted data as well as to appropriately balance ensuring access with security and privacy concerns. Requests for access to restricted or sensitive data are frequently resolved quickly.	☐ The agency possesses a central repository for the datasets they use to train AI models that allows datasets to be flexibly managed. ☐ Where feasible, the agency contributes to a publicly accessible and searchable national catalog of existing datasets that includes information about dataset usage and metadata, as well as other relevant resources. ☐ As appropriate, access policies are continually reviewed and updated to ensure that they account for privacy and security concerns.

Subdomain 2.2: Infrastructure

Definition

The agency possesses the necessary software and hardware infrastructure to develop and use AI systems.

Topics and goals

1. **Computing infrastructure.** The agency is able to provide its AI systems with the requisite secure computing capacity.

2. **Data storage infrastructure.** The agency possesses sufficient data storage resources, as well as software for maintaining stored data.

3. **Testing infrastructure.** The agency possesses AI-specific testbeds and sandboxes (software environments) that allow employees to test, develop, and experiment with AI systems.

4. **Infrastructure accessibility.** The agency's employees are able to access AI-related infrastructure as needed.

Tangible outputs

Products and processes

- Physical or cloud-based secure computing systems designed to support AI systems

- Databases, data warehouses, and file systems designed for AI-related use cases

- AI sandboxes

- AI testbeds

- A system for requesting access to AI-related infrastructure

Subdomain 2.2: Infrastructure

The agency possesses the necessary software and hardware infrastructure to develop and use AI systems.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
2.2.1 Computing infrastructure	☐ The agency does not have access to computing infrastructure designed for and dedicated to managing AI computing loads.	☐ The agency can access some AI-dedicated computing infrastructure, but it is either insufficient to meet the agency's needs or unavailable on a whole-of-agency level.	☐ The agency provides sufficient computing capacity to support its AI systems. ☐ The agency employs hardware and software strategies (such as parallel processing) to enhance computing speed, efficiency, and performance of AI systems.	☐ The agency continuously and systematically reviews and updates both software and hardware computing infrastructure to ensure that the agency's AI computing needs are met.
2.2.2 Data storage infrastructure	☐ The agency does not have access to data storage infrastructure designed for and dedicated to addressing the data storage needs of AI systems.	☐ The agency has access to some AI-dedicated data storage, but it may be insufficient or unavailable on a whole-of-agency level.	☐ Agency AI developers either possess or can access secure databases, data warehouses, and file systems designed for AI-related use cases. ☐ The agency possesses software designed to prepare, clean, and maintain the data necessary for AI systems to operate.	☐ The agency continuously and systematically reviews and updates both software and hardware data storage infrastructure to ensure that the agency's data storage needs are met. ☐ The agency's data and computing infrastructure are colocated.
2.2.3 Testing infrastructure	☐ The agency does not have access to sandboxes or other testing environments for AI experimentation.	☐ The agency can access some form of AI sandbox or other AI testing environment, but it is either insufficient to meet the agency's needs or not available on a whole-of-agency level.	☐ The agency operates sandboxes or other testing environments and makes them available to employees for safe and ethical experimentation with using AI models.	☐ The agency continuously and systematically reviews its testing infrastructure to make sure that the agency's needs are met.

2.2.3 Testing infrastructure			☐ The agency operates or has access to AI-specific testbeds with the hardware, software, and computing capacity necessary to test and develop AI systems.	
2.2.4 Infrastructure accessibility	☐ The agency does not have formalized processes or structures in place to help employees access the AI infrastructure they need.	☐ The agency may have a formalized process in place to help employees access AI infrastructure, but it is incomplete, not well known, and/or difficult to use.	☐ The agency has a system that allows employees to request computing or testing time on the agency's infrastructure. The system is used regularly, and infrastructure access is typically not a major challenge.	☐ The agency's software and hardware infrastructure enable interagency sharing of computing, data storage, and testing infrastructure when possible.

Subdomain 2.3: Procurement

Definition

The procurement of AI systems is systematized

Topics and goals

Procurement processes. The agency procures AI-related assets in a systematic way through a consistent and coordinated process.

Tangible outputs

Policies, plans, and guidance

- An AI procurement strategy

- Standard acquisition language and guidance for AI technology acquisition

Subdomain 2.3: Procurement

The procurement of AI systems is systematized.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
2.3.1 Procurement processes	☐ Templates, guidance, and best practices specific to AI-related procurements are not shared or employed throughout the agency. Offices and programs procuring AI products and services develop acquisition materials on their own.	☐ Neither centralized guidance nor support materials specific to procuring AI products and services are available, but individual offices and programs frequently share procurement language and other materials, even if driven by individual relationships. ☐ Uncoordinated AI procurements have resulted in the inefficient use of funds or technical challenges. For example, the AI systems procured by the agency's offices may be incompatible with each other or unnecessarily redundant.	☐ The agency makes standard acquisition language and guidance for AI technology available to all offices and programs. The guidance addresses vendor lock-in risks, data rights, model ownership, and intellectual property. The agency can identify numerous procurements that have used this language and guidance. ☐ The agency has a centralized procurement strategy to ensure that AI-related systems and software across the organization are compatible and not unnecessarily redundant.	☐ The agency collaborates with interagency and industry partners to determine AI procurement best practices and to ensure that procurement policies provide proper oversight of AI systems. ☐ The agency works with its interagency partners to standardize procurement strategies so that AI-related systems and software are compatible between different agencies.

Subdomain 2.4: Workforce and expertise

Definition

The agency is able to recruit, retain, and develop personnel with AI expertise as well as foster an AI-literate workforce.

Topics and goals

1. **Skills and workforce planning.** The agency understands the current state of its AI workforce and can identify the AI-related skills it needs, prioritize gaps in those skills, and forecast future AI-related skill needs.

2. **Recruitment and retention.** The agency is able to attract, hire, and retain individuals with the AI-related skillsets that the agency needs.

3. **Fostering of emerging talent.** The agency provides its employees with opportunities to build and enhance their AI skillsets.

4. **Enhancement of AI literacy.** The agency educates its employees on how to apply AI technologies to their work, as well as the legal, ethical, and policy implications of AI use.

Tangible outputs

Policies, plans, and guidance
- An AI skills framework

People and organizations
- AI mentorship programs and communities of interest

Products and processes
- A diagnostic assessment of the agency workforce's current and needed AI skills

- A suite of incentives used to attract AI professionals

- AI-related training materials for employees

- A central repository for authoritative guidance on AI use

Subdomain 2.4: Workforce and expertise

The agency is able to recruit, retain, and develop personnel with AI expertise as well as foster an AI-literate workforce.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
2.4.1 Skills and workforce planning	☐ The agency has not attempted to systematically identify the AI-related skills that are needed throughout its workforce. ☐ The agency does not incorporate its AI workforce needs into overarching strategic workforce planning efforts.	☐ Some agency offices have begun to identify the AI skills needed in their workforces, but these efforts are not coordinated or the findings are not disseminated on a whole-of-agency level. ☐ The agency has made efforts to incorporate its AI workforce needs into its strategic workforce planning efforts, but these efforts have led to few tangible outcomes to date.	☐ The agency has conducted a diagnostic AI skills assessment to improve its understanding of its workforce's current AI skills. The assessment addresses skills related to the intersection of AI with other fields (e.g., the legal implications of AI use) in addition to purely technical skill areas. ☐ The agency has a framework that identifies necessary AI skills, prioritizes current skill gaps, and forecasts future skills needs. The agency can list specific skill areas it has prioritized with respect to recruitment and development of its AI workforce. ☐ The agency's AI workforce needs are incorporated into the agency's strategic workforce planning and associated budgets.	☐ The agency collaborates on its AI hiring strategy with interagency partners, including by sharing best practices and collectively identifying key skill gaps it needs to fill.
2.4.2 Recruitment and retention	☐ Offices within the agency use minimal coordination when hiring individuals with AI-related expertise.	☐ While some subcomponents of the agency offer incentives to attract AI professionals, this approach is not systematically applied throughout the agency or has not proven to be effective.	☐ Offices within the agency coordinate on hiring individuals with AI expertise in order to address skill gaps and avoid excess capacity.	☐ The agency regularly identifies and circulates best practices for hiring, retaining, and empowering individuals with AI expertise.

2.4.2 Recruitment and retention	☐ The agency does not use incentives or other programs to attract AI expertise or does so very infrequently. ☐ The agency cannot quantify its needs for staff with AI skills, or it has extensive unmet needs for staff with such skills.	☐ Although recruitment and retention levels have begun to meet staffing needs, significant gaps remain.	☐ To attract and retain skilled AI professionals, the agency uses incentives, pay-setting authorities, and other compensation flexibilities similar to those available for cyber and information technology positions. ☐ Most agency positions requiring staff with AI skills are filled; staffing levels are roughly in line with staffing needs.	
2.4.3 Fostering of emerging talent	☐ The agency does not provide professional development opportunities for employees seeking to develop AI-related expertise or does so very infrequently.	☐ Opportunities to work on AI-related projects are not frequently provided to employees and largely depend on personal connections or happenstance. ☐ There are some mentorship structures in place for AI professionals, but they are siloed across the agency or informal in nature.	☐ The agency broadly communicates opportunities to work with AI systems to employees. Announcements of such opportunities occur regularly. ☐ The agency hosts mentorship programs and communities of interest to share best practices and connect established AI experts with emerging AI professionals. ☐ The agency provides pathways for individuals with nontraditional academic, professional, or disciplinary backgrounds to access AI-related roles.	☐ The agency participates in interagency AI mentorship programs or communities of interest, or similar programs that include external partners.

| 2.4.4 Enhancement of AI literacy | ☐ The agency does not encourage or offer formal training to its employees on AI best practices, policies, and use cases, or it does so very infrequently.

☐ The agency's employees frequently report being unsure of how to apply AI technologies to their own work. | ☐ The agency provides employees with some trainings on AI, but these trainings may only be provided to certain offices or the content may be limited.

☐ Employee trainings on AI are not designed to meet specific agency needs.

☐ Guidance or training materials on AI are not stored in a central repository, meaning that individual offices (and agency leadership) may not be fully aware of guidance and trainings developed or used by other offices. | ☐ The agency provides all appropriate employees with trainings that give them a foundational understanding of AI, including the legal, ethical, and policy considerations of AI implementation as well as its technical aspects.

☐ The agency trains appropriate employees on how to apply AI technologies to their work.

☐ A central repository for authoritative guidance on AI knowledge is available. | ☐ Best practices and materials for AI trainings are shared at the interagency level.

☐ Trainings are continuously adapted as AI technologies develop.

☐ The agency has processes in place to ensure that employees take or repeat AI trainings on a regular basis. |

Domain 3: Outputs are impactful

Domain 3: Outputs are impactful	
The agency has established performance standards, testing procedures, and policies to ensure that AI systems are applied to suitable use cases and contribute to the agency's overall mission.	
Subdomain	**Definition**
Performance	The agency tests, monitors, and refines the performance of its AI systems.
Use cases	The agency can identify and apply AI systems to beneficial use cases.

Subdomain 3.1: Performance

Definition
The agency tests, monitors, and refines the performance of its AI systems.

Topics and goals
1. **Performance indicators and metrics.** The agency has clearly defined and systematically employed a set of robust performance indicators and metrics that are paired with rigorous testing methodologies.

2. **Testing.** The agency uses rigorous testing methodologies supported by clearly defined procedures that stipulate how and when to carry out performance assessments of its AI systems both before and during deployment. The agency has also defined actions that can be taken if a system ceases to perform properly.

3. **Monitoring.** The agency has defined procedures for identifying and monitoring AI systems that change over time, as well as for identifying and implementing necessary course corrections.

4. **Feedback and optimization.** The agency has a systematized process for identifying failures and using them to inform and improve future AI development efforts.

Tangible outputs
Policies, plans, and guidance
- Guidance on how and when to assess AI system performance

- Policies to address performance shortfalls and other issues with AI systems

Products and processes
- A set of indicators and metrics that can be used to assess AI system performance

- Test sets and methodologies to measure the agency's performance indicators and metrics

- A set of quantitative baseline performance targets and acceptable ranges of drift from that performance baseline

- A process to identify and incorporate AI system failures into lessons learned

Subdomain 3.1: Performance

The agency tests, monitors, and refines the performance of its AI systems.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
3.1.1 Performance indicators and metrics	☐ AI system performance is judged on an ad hoc basis. The agency lacks metrics for evaluating the validity, efficiency, explainability, interpretability, and reliability of its AI systems or their outputs. ☐ The agency has not defined test methodologies or test datasets that it uses to measure AI system performance.	☐ Efforts to develop AI performance indicators are nascent. For example, many of the agency's performance indicators are limited in scope, use limited measurement methods, do not address system goals and objectives, or do not address human-AI teaming dimensions. ☐ Performance evaluation methodologies and test datasets are defined for some agency AI systems.	☐ The agency has clear AI metrics that assess AI system performance for properties such as explainability, interpretability, reliability, validity (e.g., false positive rate, false negative rate), and efficiency (e.g., training times, prediction latency). ☐ AI system performance metrics frequently use a variety of inputs, including end user and community feedback, inputs from system operators, and quantitative measures of AI system outputs. Metrics consistently take into account system requirements and measure progress toward program goals and objectives. Metrics generally address human-AI teaming dimensions when appropriate. ☐ Performance metrics are paired with clearly defined test datasets and detail the evaluation methodologies for the vast majority of the agency's AI systems.	☐ Performance indicators are context-specific, are designed to enable continuous improvement, and incorporate contributions to wider organizational goals. ☐ The agency periodically reviews its measures, metrics, and testing procedures to ensure that they are an effective means of measuring AI system performance as technologies and use cases change.

| 3.1.2 Testing | ☐ The agency has no central guidance on how and when to carry out AI system performance assessments; in addition, it has not identified the intervals during which AI system performance should be assessed.

☐ Substantial numbers of AI systems do not undergo performance testing or validation before deployment or during operations. | ☐ The agency has issued guidance for how and when to assess AI system performance, but this guidance is inconsistently applied in practice.

☐ AI systems are sometimes tested for performance and validated before deployment, but such testing occurs inconsistently or does not occur under realistic operational conditions. | ☐ The agency has documented guidance and standards for how and when to assess AI system performance, including guidelines on experimental design, data quality, model training, system testing, and system validation. Guidance and standards address the appropriate intervals for assessing AI system performance.

☐ AI systems and human-AI configurations are tested for performance and validated prior to deployment. Performance testing considers progress toward stated goals and occurs under a wide range of close to real-world operational conditions. A rigorous software testing regimen that includes comparison to performance benchmarks and formalized reporting and documentation of results exists. Testing includes the use of pilot programs, red-team testing, and testing in sandbox environments. | ☐ The agency has procedures to evaluate, document, and improve guidance and standards for AI system performance testing as systems evolve and their use cases change.

☐ The agency periodically reevaluates and revises its performance testing approaches and regimen to address changing technologies and use cases more effectively. |

3.1.2 Testing		☐ Some performance testing may be carried out after deployment to discern any differences between pre- and postdeployment performance, but these tests are irregular or do not reliably lead to system updates when issues are identified.	☐ AI systems undergo regular performance testing while in operation. The agency regularly compares predeployment system testing results to postdeployment system performance. When differences are observed, the agency identifies causes and updates testing methods accordingly.	
3.1.3 Monitoring	☐ The agency has not identified AI systems that change (in their performance, nature, or other characteristics) over time, such as adaptive or online learning AI models. ☐ The agency does not monitor its systems for model and data drift, or have mechanisms for correcting drift. The agency does not meaningfully monitor the data input into its AI systems or their outputs.	☐ The range of acceptable data and model drift is established for some systems. The agency sometimes monitors for drift from target performance range, but targets may be qualitative and loosely defined, and drift out of target range may not result in corrective actions on the part of the agency. ☐ The agency monitors trends in input data distribution and generated predictions, but monitoring may be inconsistent or ad hoc, and significant variance in these trends may not trigger alerts or resultant action.	☐ The agency has identified its AI systems with underlying models that evolve over time and regularly assesses them for unforeseen changes in behavior or performance. ☐ The agency defines quantitative targets for baseline system performance and acceptable ranges of drift from baseline performance; it also consistently monitors system performance relative to these ranges and takes corrective actions when system performance deviates from the acceptable ranges.	☐ Ongoing monitoring of AI systems is well defined and consistently applied across systems; in addition, there are procedures in place to continually refine, improve, and clarify these monitoring efforts. ☐ Processes for identifying and implementing course corrections are periodically reviewed, critically analyzed, and altered as necessary in a process of continuous improvement.

3.1.3 Monitoring	☐ The agency does not have plans that define the actions that should be taken to address issues with AI systems, or for shutting down AI systems in the case of malfunction	☐ The agency has plans that define actions to be taken to address system issues and for shutting down systems in case of their malfunction for some of its AI systems. Plans may be undertested or otherwise lack rigor.	☐ The agency has policies in place that outline specific actions that should be taken to address issues with AI systems. The agency has a robust capability that allows it to temporarily shut down AI systems if necessary.	
3.1.4 Feedback and optimization	☐ The agency does not identify modes of failure for AI systems or does so very infrequently. There is little or no attempt to incorporate, correct, or otherwise optimize AI systems according to failures that occur. ☐ Results that deviate from desired outcomes largely fail to inform future efforts to develop and deploy AI systems. ☐ The agency has not established methods for users or parties affected by AI system outputs to provide feedback at any point in the system's life cycle. Any feedback from users or affected parties is obtained through ad hoc, unofficial, or nonspecialized avenues.	☐ Modes of failure have been discussed but not defined or consistently and reliably identified. There is little or no attempt to correct or otherwise optimize AI systems according to observed failures. ☐ Results that deviate from desired outcomes are incorporated as lessons learned in future AI-related efforts in an ad hoc or informal manner on an office-by-office, system-by-system, or project-by-project basis. ☐ The agency has methods for users or parties affected by the outputs of the AI system to provide feedback on the system, but these methods may be poorly communicated or otherwise unideal, resulting in little actionable feedback being obtained.	☐ Modes of failure have been identified and defined. Observed failures inform efforts to optimize AI systems. ☐ The agency has an established process through which results that deviate from desired outcomes can be incorporated as lessons learned in future AI-related efforts. There are mechanisms in place that enable individuals from different offices within the agency to share lessons learned. ☐ The agency has established means for users or parties affected by the outputs of the AI system to provide feedback on the system throughout its life cycle.	☐ The agency routinely refines the process through which it identifies modes of failure and uses observed failures to optimize AI systems. ☐ The agency routinely refines the process through which results that deviate from desired outcomes are incorporated into lessons learned. These lessons learned are shared throughout the agency and the interagency. ☐ The agency periodically refines the methods it employs to allow users or parties affected by AI system outputs to provide feedback.

Subdomain 3.2: Use cases

Definition
The agency can identify and apply AI systems to beneficial use cases.

Topics and goals

1. **Identification of use cases.** The agency identifies and prioritizes potential use cases to which AI systems could be applied. The agency's efforts to identify potential AI use cases draw on perspectives from around the organization and encourage employee participation.

2. **Evaluation of use cases.** The agency draws on a variety of factors and inputs to evaluate the potential use cases for its AI systems. The agency uses these evaluations to determine whether an AI system is more effective in a use case than other potential solutions and/or will provide enough benefit to justify the risks it entails.

3. **Facilitation of AI advancement in the field.** The agency actively seeks out opportunities to facilitate the development of AI capabilities in areas relevant to its mission.

Tangible outputs

Policies, plans, and guidance

- Policies that guide how agency employees should identify potential AI system use cases, and then prioritize those potential applications based on business value and technical feasibility

- A strategy identifying potential opportunities for the agency to invest in the development of AI capabilities relevant to its mission area

Products and processes

- A methodology for evaluating whether the benefits of employing an AI system outweigh potential risks

Subdomain 3.2: Use cases

The agency can identify and apply AI systems to beneficial use cases.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
3.2.1 Identification of use cases	☐ The agency has made few coordinated efforts to identify potentially beneficial use cases for AI development and deployment. Exploration of potential AI use cases is carried out informally within individual project teams. ☐ The agency and its offices make little or no attempt to apply AI systems to diverse use cases or to systematically explore potential applications of AI.	☐ Exploration of potential AI use cases is formalized but insular, often occurring without input from appropriate domain area experts or the involvement of other talent within the agency. ☐ The agency has identified potentially beneficial use cases for AI, but its process for doing so may lack rigor, be largely informal, or remain otherwise underdeveloped. Use cases that are identified are not prioritized based on their relative potential benefits.	☐ Policies and frameworks for identifying—and then prioritizing—potentially beneficial use cases are well established and consistently applied. The agency has identified its most significant potentially beneficial AI use cases and prioritized them for pilot development based on both business value and technical feasibility. The needs of these use cases are clearly defined, to avoid wasting resources on developing or acquiring unnecessary capabilities. ☐ Exploration of potential AI use cases is informed by both technical and domain-area expertise. Employees from around the agency are encouraged to take part in exploring potential AI use cases.	☐ The agency regularly identifies, prioritizes, and develops beneficial AI applications. This process happens according to a codified and coherent procedure, which is itself periodically evaluated for performance and improvement. ☐ Relevant expertise is applied from around the organization in collaborative contexts to identify and explore potential use cases. In addition, lessons learned from effective and ineffective use cases—as well as those learned about the generalizability of different AI systems—are documented and incorporated into future efforts to identify possible AI use cases.

3.2.1 Identification of use cases		☐ The agency may make some attempts to apply AI systems to multiple, diverse use cases, but these efforts may be only semiformalized or otherwise less than optimal; as a result, the agency is still applying AI systems to a relatively homogenous set of use cases.	☐ The agency has employed AI across a broad range of applicable operational and programmatic use cases.	☐ The agency actively seeks out new application areas. A framework for developing, expanding, and scaling current systems for application to new use cases is consistently applied, and both positive and negative outcomes are incorporated into future development efforts.
3.2.2 Evaluation of use cases	☐ The agency does not routinely or systematically evaluate use cases for business value and technical feasibility. ☐ When the agency does evaluate potential AI use cases, it does so in an ad hoc manner and may not make an effort to test the efficacy of an AI model within a use case.	☐ The agency evaluates potentially beneficial AI use cases for business value and technical feasibility, but these evaluations do not generally follow a standard set of procedures. ☐ Potential use cases are evaluated according to limited and narrowly defined measures. Minimum viable products, proof-of-concept systems, and pilot projects have been developed to demonstrate the efficacy of AI models in some use cases, but this is not done systematically.	☐ Evaluation of use cases for business value and technical feasibility occurs through established and coherent procedures and takes place throughout the life cycle of AI systems, including before they are developed and implemented. ☐ Potential use cases are evaluated using a wide range of factors in addition to desired outputs and system performance. These factors can include data availability, stakeholder inputs, and expert interviews, as well as the systematic employment of minimum viable products, proofs-of-concept, and pilot projects.	☐ Use cases are evaluated for business value and technical feasibility using a diverse set of criteria and input streams. The process and procedures used to evaluate use cases are defined and supported by an institutional framework. These procedures are periodically reconsidered, with opportunities for improvement identified and the overall process streamlined. ☐ In addition to systematically assessing whether the benefits of employing certain AI systems outweigh the potential risks or drawbacks, the agency also regularly reviews the process and methodology through which it makes those assessments.

| 3.2.2 Evaluation of use cases | ☐ The agency does not assess whether the benefits of employing a given AI system outweigh the potential risks or drawbacks or does so very infrequently.

☐ The agency does not assess whether employing an AI-enabled solution in a given use case provides a comparative advantage over traditional solutions or does so very infrequently. | ☐ The agency may assess whether the benefits of employing certain AI systems outweigh the potential risks or drawbacks, but it may do so inconsistently, or these efforts may lack top-down guidance.

☐ The agency may assess whether using an AI-enabled solution in a given use case provides a comparative advantage over traditional solutions, but it may do so inconsistently or without top-down guidance. | ☐ The agency uses a consistent methodology to systematically assess whether the benefits of employing AI systems outweigh the potential risks or drawbacks. Potential risks and impacts are evaluated in the context of likelihood and magnitude using a systematic process for prioritization and management.

☐ When planning AI systems, the agency consistently examines whether AI systems will provide functions more effectively, efficiently, economically, and ethically relative to human baselines, conventional approaches, or the status quo. The agency understands AI capabilities, targeted usage, goals, and expected benefits and costs compared with the status quo and can quantify them. | ☐ The agency regularly reviews whether the AI systems it uses continue to function more effectively, efficiently, economically, and ethically than conventional approaches, newly developed AI systems, or other solutions. |
| **3.2.3 Facilitation of AI advancement in the field** | ☐ The agency has not considered its role in advancing AI capabilities relevant to its mission area. | ☐ The agency has identified potential roles it could play in facilitating the advancement of AI capabilities relevant to its mission area, but it has not yet taken concrete actions to fulfill such roles. | ☐ The agency supports and invests in cyberinfrastructure, education and training, data, and computational power to facilitate the advancement of AI capabilities relevant to its mission area. The agency follows a set of consistent procedures for identifying new opportunities for providing this support. | ☐ The agency actively seeks out opportunities to support and invest in expanding AI capabilities relevant to its mission area. The agency also regularly reexamines its role in providing support and investment for AI capabilities relevant to its mission area, evaluating prior efforts for efficacy and areas of improvement. |

Domain 4: Products and results are trustworthy

Domain 4: Products and results are trustworthy The agency designs and manages AI systems so that they are representative, transparent, and unbiased.	
Subdomain	**Definition**
Representative	Affected stakeholders are able to participate in AI system development, and individuals with varying abilities and backgrounds can use AI systems.
Transparent	The development and use of AI systems is thoroughly documented, and AI system operations and results are readily understood.
Unbiased	The AI system is designed and managed in such a way as to reduce the prevalence of biased outcomes.

Subdomain 4.1: Representative

Definition

Affected stakeholders are able to participate in AI system development, and individuals with varying abilities and backgrounds can use AI systems.

Topics and goals

1. **Engagement with external stakeholders.** The agency consults with the communities affected by the AI systems it uses as well as community-based organizations, civil liberties organizations, and other external stakeholders, and it is able to use inputs derived from these engagements to reduce bias and discrimination in the AI systems it uses.

2. **Participation of internal offices that represent stakeholders.** The agency draws on internal offices that represent affected stakeholders, such as civil rights and civil liberties offices, to support its efforts to reduce bias and discrimination in the AI systems it uses.

3. **User experience.** The agency designs the interfaces of its AI systems to be accessible to individuals with disabilities or nontechnical backgrounds.

Tangible outputs

Policies, plans, and guidance
- Policies describing how and when the agency should consult with the public about AI systems

- Guidance on how agency employees and offices should engage external stakeholders about AI systems issues

Products and processes
- A set of design and implementation principles for AI system interfaces

Subdomain 4.1: Representative

Affected stakeholders are able to participate in AI system development, and individuals with varying abilities and backgrounds can use AI systems.

Topic	Stage of Maturity			
	Developing	**Performed**	**Established**	**Optimized**
4.1.1 Engagement with external stakeholders	☐ The agency has not meaningfully consulted with the communities that will be affected by the AI systems the agency employs. ☐ Little or no engagement with communities or stakeholders that have an interest in reducing bias and discrimination in AI systems has occurred across the agency.	☐ Consultation on AI systems with affected communities occurs sporadically and is driven by the individual project; the consultation is also frequently limited to specific phases of the life cycle of AI systems or in response to specific problems or issues. ☐ Some offices and programs engage with communities or stakeholders that have an interest in reducing bias and discrimination in AI systems, but this engagement is inconsistent across the organization.	☐ The agency regularly consults affected communities throughout the life cycle of its AI systems that affect the public, and it has policies guiding when such consultation is required and how it should be carried out. These consultations regularly provide user-driven requirements, advice on AI technologies, and suggestions for reducing system bias. ☐ Agency offices regularly engage with community-based organizations, civil rights and civil liberties organizations, academic institutions, industry, and state, local, territorial, tribal, and federal government agencies in efforts to reduce bias and discrimination in AI systems.	☐ The agency can identify numerous examples in which external consultations have affected AI system design or implementation substantially. ☐ The agency can identify several cases in which input from community engagements, civil rights offices, or civil liberties offices has affected the design, deployment, and use of its AI systems.

4.1.1 Engagement with external stakeholders	☐ The agency has not established standard methods or best practices for engaging with external stakeholders.	☐ Standard methods or best practices for engaging with external stakeholders are not shared throughout the agency. Offices and programs that choose to engage with external stakeholders on AI issues are left to develop their own engagement techniques.	☐ Agency offices regularly engage with community-based organizations, civil rights and civil liberties organizations, academic institutions, industry, and state, local, territorial, tribal, and federal government agencies in efforts to reduce bias and discrimination in AI systems. ☐ The agency provides central guidance and best practices for engaging external stakeholders on AI systems issues and using their input. Offices and programs regularly consult such guidance and have reported it to be effective.	☐ The agency can identify specific efforts it has made to engage with the entire population affected by its AI systems, including underrepresented, vulnerable, and protected demographics.
4.1.2 Participation of internal offices that represent stakeholders.	☐ Offices that represent AI system stakeholders, such as civil rights and civil liberties offices, generally do not participate in the planning or monitoring of AI systems.	☐ Offices that represent stakeholders, such as civil rights and civil liberties offices, sometimes participate in the planning or monitoring of AI systems.	☐ Offices that represent stakeholders, such as civil rights and civil liberties offices, consistently participate in the planning and monitoring of AI systems.	☐ The agency can identify numerous examples in which input from offices that represent stakeholders, such as civil rights and civil liberties offices, have influenced how AI systems are developed and used.
4.1.3 User experience	☐ The agency does not regularly consider usability, human-centered design, or other aspects of user experience of its AI systems. As a result, the agency's AI systems are largely inaccessible to individuals with disabilities or those with nontechnical backgrounds.	☐ The agency has a set of design and user experience principles, but they are vague or difficult to systematically implement. As a result, many of the agency's systems may still be inaccessible to individuals with disabilities or nontechnical backgrounds.	☐ The agency has a well-developed set of design and implementation principles focused on user experience, with appropriate functions such as collaboration tools, search tools, user guides, and other tools and functions where warranted.	☐ The agency's design, implementation, and user experience principles are well developed and thorough, and they are expressed in consistent standards that employees can meet. There is an explicit procedure to evaluate, improve, and reevaluate these standards, as well as functions included in the user experience.

4.1.3 User experience			☐ The agency has considered the needs of users with varying sets of abilities, skills, and backgrounds, and has integrated pathways and tools for learning into its user experience. As a result, the agency's AI systems are accessible to individuals with disabilities or nontechnical backgrounds.	☐ The agency actively solicits feedback from a breadth of users in an effort to continuously improve and update tools designed to help new users learn how to operate the system.

Subdomain 4.2: Transparent

Definition

The development and use of AI systems is thoroughly documented, and AI system operations and results are readily understood.

Topics and goals

1. **Descriptive documentation of AI systems.** The agency has documented the goals, use cases, model structures, risks, and other qualities of its AI systems, and it makes this documentation widely available.

2. **Identification of AI content.** The agency requires employees and offices to mark content created by generative AI systems and consistently alerts users when they are interacting with an AI system.

3. **Development of transparent systems.** The agency designs its AI systems with features that enhance transparency and enable the agency to trace the data used by an AI system to generate a specific output.

4. **Data transparency.** The agency generates and makes available metadata describing the data the agency used to train its AI systems.

5. **Communication of AI system performance.** The agency releases information about AI system performance at appropriate intervals.

6. **Third-party systems.** The agency has oversight of the AI systems it obtains from third-party providers.

Tangible outputs

Policies, plans, and guidance

- Policies mandating what aspects of AI systems should be documented

- Policies on how offices and employees should identify and mark content produced by generative AI systems

- A framework describing how, when, and which AI system performance information should be released to the public

Products and processes

- Documentation of AI system goals, use cases, requirements, designs, model structures, operations, evolutions, risks, and limitations

- A portal or other service that provides access to searchable AI system documentation

- AI systems that incorporate transparency features, such as audit logs and histories

- Metadata that document the provenance of data used to train AI systems, as well as how those data were prepared for input into an AI system

- A standardized set of documentation templates and other resources for use in third-party AI system inventory and approval

Subdomain 4.2: Transparent

The development and use of AI systems is thoroughly documented, and AI system operations and results are readily understood.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
4.2.1 Descriptive documentation of AI systems	☐ The agency provides documentation for very few of its AI systems. Documentation that does exist describes only limited aspects of the system, and stakeholders frequently do not understand what AI systems the agency employs and the basics of how they work. ☐ Any documentation that the agency does collect is stored in a noncentralized and hard-to-navigate manner.	☐ Some agency offices and programs document their AI systems to various extents, but these activities are largely informal and not guided by policy. In addition, documentation may describe the final AI system as deployed but not the system's development and implementation. ☐ The agency maintains a catalog of AI systems, but access to its AI-related documentation remains restricted or difficult because of a lack of centralized public-access portals, unsystematic organization, or incomplete documentation.	☐ The agency has made documentation of the vast majority of its AI systems available to stakeholders, including stakeholders from outside of the agency who are impacted by the systems. Documentation covers the systems extensively, including their goals, use cases, requirements, designs, model structures, operations, evolutions, risks, and limitations. ☐ The agency has policies that require AI systems to be documented in a standardized way and to remain current. The agency shows a high level of compliance with these policies, and consumers of the documentation agree that it is clear and appropriate. ☐ The agency has disseminated available documentation to the greatest feasible extent, accounting for concerns such as cost and security. AI system documentation is inventoried and searchable, and the agency has developed portals or services that provide access.	☐ The agency has a well-defined, thorough, and easy-to-use recordkeeping system for AI materials. This system incorporates internal feedback in a process of continuous improvement.

4.2.1 Descriptive documentation of AI systems			☐ The agency systematically archives AI-related documentation.	
4.2.2 Identification of AI content	☐ The agency frequently provides AI-generated content to consumers without identifying it as such. ☐ Users often interact with agency AI systems—or AI-enabled systems—without being aware that they are doing so.	☐ Many agency offices and programs identify the AI-generated content they provide, but this is performed inconsistently across the agency. ☐ The agency informs users on an ad hoc basis of when they are interacting with an AI system or an AI-enabled system.	☐ Agency policy requires offices and programs to identify content created by generative AI systems via watermarking or other labeling of synthetic content. Offices and programs consistently comply with this policy. ☐ The agency consistently alerts users when they are interacting with an AI system or an AI-enabled system.	☐ The agency periodically reviews the systems and methods it uses to inform users of when they are interacting with an AI system or AI-generated content. These reviews are used to improve and optimize those systems and methods.
4.2.3 Development of transparent systems	☐ Very few of the agency's AI systems include features or systems that allow a user or affected party to understand or trace the system's decisions and outputs. Many models are considered to be "black boxes."	☐ Some agency systems have features such as audit logs that support post hoc understanding of their outputs. ☐ Results of AI systems are sometimes—but not consistently—traceable.	☐ Many agency systems are instrumented for measurement and tracking as they are developed. Systems frequently maintain histories, audit logs, and other information that can be used by AI actors to review outputs and evaluate possible sources of error, bias, or vulnerability. ☐ The data or knowledge that AI systems use to produce outputs and make decisions can frequently be identified and are consistently and easily traceable.	☐ The agency periodically reviews and revises the guidance it provides to AI system developers on improving the traceability and transparency of AI system decisions and outputs.

4.2.4 Data transparency	☐ The agency does not have established processes or mechanisms for tracing the provenance of data used to train its AI systems.	☐ There may be some documentation of data provenance, but that documentation is not systematically recorded or made readily available.	☐ The agency collects and makes readily available metadata that document the provenance of data used to train AI systems, including sources, origins, transformations, augmentations, labels, dependencies, and constraints, as well as the rationale behind how developers organized, revised, and segregated their data in preparation for model training, development, and testing.	☐ When possible, the agency makes the metadata and documentation they produce available to all users of their datasets and the public.
4.2.5 Communication of AI system performance	☐ AI system performance information is infrequently or never shared with stakeholders or affected communities.	☐ Although some AI system performance information is shared with stakeholders and those affected by system decisions, such information sharing is performed irregularly and inconsistently across the agency.	☐ The agency has a framework guiding the release of AI system performance information to stakeholders and affected communities. Many agency offices and programs have used the framework to establish the frequency and scope for sharing metrics and related information about AI system performance.	☐ Third parties have independently assessed the performance of agency AI systems, and the agency has made the results available to stakeholders and affected communities. ☐ Where possible, AI system stakeholders and affected communities have been allowed to test AI systems.
4.2.6 Third-party systems	☐ The agency generally has very little insight into the operation of third-party AI systems it acquires, and its ability to evaluate them is highly limited.	☐ The agency has some requirements for third-party AI system documentation, although the requirements are limited in their extent.	☐ The agency places clear requirements on third-party AI system providers for describing the operations of their systems, including requirements for documenting the data and algorithms used, assumptions, and limitations.	☐ The agency's framework for supplying information to and communicating with third parties is shaped by collaboration and feedback with those third parties in a process of continuous improvement.

4.2.6 Third-party systems		☐ The agency has developed documentation to help users and third parties ensure that the AI system is used as intended. ☐ The agency has a limited capability to independently evaluate vendor-supplied systems.	☐ The agency has developed standard resources such as model documentation templates and software safelists to assist in third-party technology inventory and approval activities. ☐ The agency is capable of performing independent evaluations on vendor-supplied AI systems.	

Subdomain 4.3: Unbiased

Definition

The AI system is designed and managed in such a way as to reduce the prevalence of biased outcomes.

Topics and goals

1. **Characterization of bias and discrimination.** The agency can identify negative effects AI systems might have on subsets of the population and understands the protected attributes that cannot be used by AI systems for decision-making.

2. **Metrics for evaluating bias.** The agency can measure AI system bias in a consistent manner and has defined acceptable levels of differences in system performance across different groups.

3. **Monitoring for biased outcomes.** The agency monitors its AI systems for instances of bias and tracks incidents in which AI systems produce unjust results.

4. **Unbiased training data.** The agency tests its training data for biases and representativeness and tests trained AI models for emergent biases.

Tangible outputs

Policies, plans, and guidance

- Guidance that identifies subsets of the population who could be harmed by AI system bias

- A framework and metrics for measuring AI system bias and discrimination, as well as defined levels of bias that are (and are not) acceptable

- Guidance on how to ensure that AI training data are representative

Products and processes

- Mechanisms for AI system stakeholders to report and provide feedback on AI system bias and discrimination

Subdomain 4.3: Unbiased

The AI system is designed and managed in such a way as to reduce the prevalence of biased outcomes.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
4.3.1 Characterization of bias and discrimination	☐ Agency offices and programs rarely or never identify the ways systems could negatively affect individuals or groups as part of the AI system planning process. ☐ Agency offices and programs rarely or never identify the subsets of the population that could be negatively affected by their AI systems. ☐ The agency does not have any policy governing protected attributes that AI systems cannot use for predictive or decision-making purposes.	☐ Although several agency offices and programs have identified potential negative effects of AI systems on individuals or groups as part of system planning, this is inconsistent across the agency. There is no requirement to assess potential negative effects on individuals or groups, or the requirement is not regularly enforced. ☐ Although some agency offices and programs identify subsets of the population with which to test their systems for biased or discriminatory effects, this is left to the discretion of the office or program and not driven by policy. ☐ Agency policies address protected attributes that AI systems cannot use, but compliance with such policies is inconsistent.	☐ The agency requires that offices and programs identify the different types of negative effects that their systems could have on affected individuals or groups as part of system planning. Offices and programs consistently comply with the requirement. ☐ The agency has provided central guidance identifying specific subsets of the population that may be harmed by AI system bias and requires offices and programs to ensure that their systems treat these groups fairly. ☐ The agency has identified protected attributes that cannot be used for decision-making for legal or ethical reasons. The agency can demonstrate consistent compliance with this requirement.	☐ The agency regularly reassesses its policies and practices with respect to characterizing bias and discrimination—including the types of negative effects systems can produce, the individuals and groups that can be affected, and the attributes that systems cannot use—in efforts to improve and maintain them.

4.3.2 Metrics for evaluating bias	☐ The agency has not developed consistent metrics or methodologies for measuring bias and discrimination in AI performance. Offices and programs do not quantify bias within their systems, or they use limited or highly varied methods to do so. ☐ Acceptable levels of variance in system performance for different affected groups are defined for very few or none of the agency's systems.	☐ The agency has developed some guidance on ways to measure AI system bias. Quantitative metrics for bias are limited and inconsistently implemented. ☐ Acceptable levels of variation for system performance across different affected groups have been defined for some agency AI systems.	☐ The agency uses a consistent framework for measuring AI system bias and discrimination. The agency has developed standard metrics for bias, and individual organizations also use custom, context-specific metrics. ☐ The agency has consistently defined acceptable levels of difference in performance of its systems across different groups, along with the actions to be taken if performance differences exceed acceptable levels.	☐ AI system performance metrics that address bias and discrimination are periodically updated based on observed outcomes, reported instances of problematic results, and in-depth studies of system performance.
4.3.3 Monitoring for biased outcomes	☐ The agency has not investigated the extent to which its AI systems produce fair and consistent results for different subsets of the populations affected by them, or has done so only for a very small number of systems. ☐ The agency has not established mechanisms for those affected by AI systems to provide feedback on perceived bias and discrimination.	☐ The agency has studied the extent to which systems produce fair and consistent results for different populations for some systems but has not done so comprehensively. ☐ The agency provides mechanisms for reporting biased or discriminatory AI results, but they are not well known or receive very little use.	☐ The agency regularly tracks the extent to which its AI systems produce fair and consistent results for different subsets of the population affected by them, including historically underrepresented, vulnerable, and protected demographics. Such checks for bias frequently incorporate downstream tasks to account for system performance in real-world deployments. ☐ AI system stakeholders have regularly used established mechanisms for providing feedback on system bias and discrimination or reporting incidents of bias or discrimination.	☐ The agency reviews the AI models it uses over a routine interval to ensure that they still operate accurately and without bias. ☐ The agency can identify several cases in which reports of incidents of bias or discrimination have affected the design, deployment, and use of its AI systems.

4.3.3 Monitoring for biased outcomes	☐ The agency is rarely or never able to identify reported or detected unjust results, misuses of AI, or AI controversies across its AI systems.	☐ The agency tracks incidents of unjust results, misuses of AI, and AI controversies for some systems, but this is done inconsistently and at the discretion of the individual office or program.	☐ The agency tracks incidents of unjust results, misuses of AI, and AI controversies for nearly all of its systems, and it shares information about such incidents with stakeholders.	☐ The agency can identify cases in which it has updated AI models to address dissimilar performances for different populations and reported or detected unjust AI results.
4.3.4 Unbiased training data	☐ There are no agency-wide procedures in place to test for biases in training data. ☐ The agency does not provide its offices with guidance on data representativeness, and very few offices and programs have assessed the training data used by their AI systems to verify that those data are representative of the range of inputs systems will receive in operational use.	☐ Some offices within the agency may test for biases in the training data they use, but there are no agency-wide procedures in place to guide these efforts. ☐ The agency has developed some guidance on verifying that training data are representative, but this guidance is limited or has not been regularly used by offices and programs that manage AI systems.	☐ Before training an AI model, the agency tests its training data for biases. Once the AI model has been trained—but before the model is deployed—the agency also tests for emergent biases that may not be apparent before training the model or cannot be discerned from looking at the model's training data alone. ☐ The agency provides guidance to its programs on ensuring that training data are representative. The guidance addresses ways to ensure that training data accurately and verifiably represent constituent populations served by the AI system with sufficient sample size and coverage of input conditions.	☐ The agency can identify cases in which evaluations of the training data used by AI systems have led to changes in AI models or their use. ☐ The agency periodically evaluates whether its guidance on ensuring data representativeness is useful and aligns with best practices, and it updates the guidance when appropriate.

Domain 5: Products are safe and secure

Domain 5: Products are safe and secure The agency effectively manages AI-related risks and protects AI systems and the data they contain from malicious attack or inappropriate disclosure.	
Subdomain	**Definition**
Cybersecurity and privacy	The agency has policies, tools, and procedures to proactively protect—and respond to cyberattacks on—its AI systems and data.
Risk management	The agency systematically identifies risks, establishes risk tolerance levels, and manages risks accordingly.

Subdomain 5.1: Cybersecurity and privacy

Definition
The agency has policies, tools, and procedures to proactively protect—and respond to cyberattacks on—its AI systems and data.

Topics and goals
1. **Mitigation of vulnerabilities.** The agency and employees occupying AI-related roles are aware of AI-specific vulnerabilities, and reviews of these vulnerabilities are incorporated into routine security processes.

2. **Assessment of safety and security impacts.** The agency performs standardized assessments of the potential safety and security risks associated with AI systems.

3. **Cyber response.** The agency understands the potential ramifications of a cyberattack on its AI systems and has implemented policies and preventative measures to address risks.

4. **Data security and privacy.** The agency can identify, categorize, and, as necessary, restrict access to data containing sensitive or private information.

Tangible outputs

Policies, plans, and guidance
- Regularly disseminated guidance on potential AI-related vulnerabilities

- Incident response plans for cyberattacks on AI systems

Products and processes
- Routine security processes that incorporate reviews for AI-related vulnerabilities

- Processes for identifying the potential safety and security effects of AI systems

- A suite of automated cyber defense tools to respond to an attack on AI systems

- Protocols that outline access controls for AI training data that include sensitive data

- Tools to trace data provenance and identify malicious data injections

- Procedures for how to inform people that their data may be used in AI system development

Subdomain 5.1: Cybersecurity and privacy

The agency has policies, tools, and procedures to proactively protect—and respond to cyberattacks on—its AI systems and data.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
5.1.1 Mitigation of vulnerabilities	☐ The agency does not review or test AI or AI-enabled systems for known AI-related vulnerabilities, or it does so very infrequently. ☐ AI developers within the agency do not regularly receive guidance on known AI-related vulnerabilities.	☐ The agency performs some monitoring against known vulnerability databases (e.g., the National Vulnerability Database), but these activities do not occur on a regular basis agencywide. ☐ The agency shares information with developers on known AI-related vulnerabilities and encourages impact assessments but does not perform oversight to ensure accountability.	☐ The agency has established consistent requirements to execute recurring reviews for known AI-related vulnerabilities as part of routine security processes. ☐ AI developers regularly receive guidance on relevant vulnerabilities from the Office of the Chief Information Officer or another centralized source.	☐ Best practices are developed for the review, identification, and resolution of any AI-related vulnerabilities identified in a cycle of continuous improvement. ☐ As allowed, the agency verifies and shares information about errors and attack patterns with incident databases, other organizations with similar systems, system users, and other relevant stakeholders. ☐ The agency performs additional analyses on AI-related vulnerabilities to generate broader insights into adversary tactics, techniques, and procedures to proactively monitor.
5.1.2 Assessment of safety and security impacts	☐ Very few of the agency's AI systems undergo safety and security impact assessments.	☐ Although safety and security impact assessments are performed for some agency AI systems, these assessments are not required, and the agency lacks a common assessment methodology.	☐ The agency has established standard processes to identify context-specific safety and security impacts associated with AI systems. AI impact assessments are required, including documentation of potential safety and security risks and their resolution.	☐ The agency examines not only risks arising from implementation of a single AI system and its direct effects, but also cascading effects and emergent effects from the agency's broader implementation of multiple AI systems.

5.1.3 Cyber response	☐ The agency has no plans or defenses in place—or very limited ones—for addressing a targeted attack on its AI systems. ☐ The agency has not identified automated cyber defense tools that could be of use during a cyberattack on an AI system. ☐ There is no systematic examination of the consequences or ramifications of different cyberattacks on an agency's AI systems.	☐ The agency is developing AI-specific incident response plans, but these plans may not be comprehensive or fully understood by implementers, or they may be incomplete. ☐ The agency has begun to identify and implement automated tools for cyber defense if a targeted attack on an AI system occurs. However, these tools may not be fully implemented or may not address an adequate range of threat vectors. ☐ The agency has some understanding of the consequences of different types of possible cyberattacks on AI systems, but this understanding has not resulted in the identification of potential failure states and alternative process redundancy.	☐ Detailed AI-specific incident response plans have been developed, including documenting the role of humans in the loop in responding to adversary attempts to disrupt AI-related operations. Plans have been practiced and tested through exercises. ☐ The agency has implemented a robust suite of automated tools for cyber defense in the event of a targeted attack on an AI system. ☐ The agency has examined and understands the ramifications of different types of possible cyberattacks on AI systems, including the potential compromising of security and privacy. The agency has identified potential failure states and alternative process redundancy as part of safety and incident response plans associated with failure of AI systems.	☐ The agency has established a cyber response that fully integrates personnel activities, automated tools, and outside assistance. ☐ Incident response plans, trainings, and exercises undergo regular cycles of review and continuous improvement to address the emergence of new threats. ☐ The agency has examined the integrity of its model training processes against sophisticated threats. ☐ The agency conducts activities such as tabletop exercises that focus on AI-enhanced attacks on critical AI systems. ☐ The agency employs various countermeasures (e.g., authentication, throttling, differential privacy, robust machine learning approaches) to increase the range of security conditions under which the system is able to return to normal function.

| 5.1.4 Data security and privacy | ☐ There are no, or very few, procedures or guidance materials in place to identify which datasets and data fields used by AI systems are sensitive.

☐ Datasets often lack clear access controls. There are no, or very few, procedures in place to protect confidential data, leaving AI system data vulnerable to unauthorized access.

☐ Low-quality data can be injected into AI system training datasets without triggering an alert to system owners. | ☐ The agency has established a consistent framework for categorizing sensitive datasets and data fields, but the framework has not yet been used to identify sensitive data across the agency's AI systems.

☐ Some data access controls are implemented. Procedures exist to protect confidential data, but they are incomplete or only partially implemented across the agency.

☐ Privacy protections on data are manually enforced and reliant on individual actors within the organization. | ☐ The agency has established a consistent framework for categorizing sensitive datasets and data fields, but the framework has not yet been used to identify sensitive data across the agency's AI systems.

☐ The agency has documented protocols and access controls for training sets or production data containing sensitive information, in accordance with privacy and data governance policies and zero-trust principles. Security and user access controls required for personally identifiable information, restricted (confidential) information, and third-party data are in place.

☐ Tools are in place to effectively and efficiently trace data provenance and facilitate the ability to ensure data integrity in the event of malicious data injection. | ☐ As necessary, the agency collaborates with privacy experts, AI end users and operators, and other domain experts to identify optimal privacy metrics for tracking within contexts of use.

☐ The agency can identify modern privacy-enhancing technologies it uses to protect its data and systems.

☐ Testing of systems for potential privacy issues takes into account the system-specific context. For example, ingestion of datasets supporting large generative AI models is done with particular caution because such datasets can memorize training data, revealing sensitive information such as home addresses or Social Security numbers in response to user prompts. |

| 5.1.4 Data security and privacy | ☐ If the agency collects human subject–related data for use in AI system development, there are no, or very few, procedures in place to inform people that their data are being collected or how their data will be collected and used. | ☐ The agency has a nascent capability to detect and respond to the injection of low-quality data into AI training data. However, this capability may be inconsistent, inefficient, or otherwise incomplete.

☐ If the agency collects human subject–related data for use in AI system development, the agency may make an effort to inform people about how their data will be collected and used, and the agency may allow them to opt out of data collection. However, this may be done inconsistently across the agency. | ☐ If the agency collects human subject–related data for use in AI system development, there are procedures in place that guide how to inform people about how their data will be collected and used as well as systems in place to allow people to opt out of data collection.

☐ The agency implements accountability-based practices in data management and protection (e.g., the OECD Privacy Principles). This includes limiting the amount of sensitive data collected or processed and ensuring that they are adequate, relevant, and not excessive to the purpose. | ☐ If the agency collects human subject–related data for use in AI system development, there are procedures in place to inform people about how their data will be collected and used and systems in place to allow people to opt out of data collection and revoke consent for the usage of data that has already been collected.

☐ The agency regularly evaluates its data access controls, including efforts to ensure that access controls do not inappropriately restrict access to data. |

Subdomain 5.2: Risk management

Definition

The agency systematically identifies risks, establishes risk tolerance levels, and manages risks accordingly.

Topics and goals

1. **Risk management approach.** The agency uses a standardized, AI-specific risk management framework to address and mitigate AI-specific risks.

2. **Risk tolerance.** The agency has defined the level of different types of risk it is willing to tolerate in its AI systems.

3. **System risk assessments.** The agency's AI users, deployers, and decision-makers understand the limitations and risks inherent in the AI systems they use, and the system's associated risk controls are centrally documented.

Tangible outputs

Policies, plans, and guidance

- An AI-specific risk management framework

- Guidance on how to prioritize the mitigation of different risks, as well as the thresholds above which different risks are no longer tolerable

People and organizations

- Roles and responsibilities for AI risk management, safety, and security

Products and processes

- Standard procedures for AI-related risk management

- A centralized repository of AI system risk controls

- System maps of the IT systems with which the agency's AI systems interact

Subdomain 5.2: Risk management

The agency systematically identifies risks, establishes risk tolerance levels, and manages risks accordingly.

	Stage of Maturity			
Topic	**Developing**	**Performed**	**Established**	**Optimized**
5.2.1 Risk management approach	☐ The agency either does not have or has a very limited AI-specific risk management framework or plan that it applies to AI projects. ☐ The agency has not defined or assigned roles and responsibilities related to AI safety and security, or it has done so to a very limited extent.	☐ Using the National Institute of Standards and Technology AI Risk Management Framework as a foundation, the agency has developed a consistent AI-specific risk management framework or plan. However, this framework or plan may not yet be widely implemented. The framework or plan may also be incomplete or may not cover emergent risks. ☐ The agency does not consistently apply its risk management framework or plan throughout the life cycle of its AI systems, and in practice it focuses on managing only limited types of risks or on addressing only those risks that arise during limited portions of an AI system's life cycle. ☐ Roles and responsibilities related to AI risk management, safety, and security may be documented, but agency staff in those roles are frequently unaware of their responsibilities or do not feel well positioned to execute them.	☐ The agency consistently uses an AI-specific risk management framework that is incorporated into the agency's broader enterprise risk management approach. ☐ The agency has established standard procedures for risk management processes, such as a standard frequency for assessing AI risks. Where appropriate, the agency adjusts procedures according to risk levels, such as monitoring high-risk systems more frequently and extensively than lower risk systems. ☐ The agency consistently applies its risk management framework or plan throughout the life cycle of AI systems and uses it to mitigate operational, technical, and societal risks.	☐ Agency personnel leverage their expertise to contribute to external guidance on secure and responsible uses of AI. ☐ The agency periodically reevaluates its risk management framework or plan for AI, including how it is being applied to AI systems, and it updates the plan or improves its application accordingly. ☐ The agency possesses the ability to extend and adapt risk management frameworks to atypical situations, such as unanticipated consequences that arise from implementation of an AI system.

5.2.1 Risk management approach			☐ Roles and responsibilities for AI risk management, safety, and security are defined and understood as they arise through various stages of the AI system's life cycle.	
5.2.2 Risk tolerance	☐ The agency has not undertaken efforts to characterize its risk tolerances or priorities with respect to AI systems risk, or it has done so to a very limited extent. ☐ The agency has not developed guidance on how its offices should prioritize mitigating different risks or identifying the threshold above which certain risks are no longer tolerable.	☐ The agency has begun efforts to characterize its tolerances and priorities for AI risks, but these efforts may be informal or unsystematic. ☐ The agency has begun to develop guidance on how its offices should prioritize risk mitigation and identify risk thresholds, but efforts are nascent.	☐ The agency has identified its tolerances and priorities with respect to AI systems risk, and they inform AI system planning throughout the system life cycle. ☐ The agency has developed and disseminated guidance on how its offices should prioritize mitigating different risks, the thresholds above which different risks are no longer tolerable, and what its offices should do if a system crosses such thresholds.	☐ The agency periodically reviews its risk tolerances and priorities to ensure that they are still relevant and reflective of the agency's goals and values. ☐ The agency regularly assesses whether its AI risk management practices align with its current risk tolerances and priorities, and it adjusts practices accordingly.
5.2.3 System risk assessments	☐ Risks associated with system or model limitations are not regularly documented for the vast majority of the agency's AI systems.	☐ Risks associated with system and model limitations are documented for some AI systems, but this is done inconsistently across the agency and at the behest of individual offices and programs.	☐ Decision-makers, deployers, and users of AI systems consistently understand the inherent uncertainties, inaccuracies, and limitations of the systems they use; the potential safety and security impacts they are exposed to; and how to use the system appropriately so as to mitigate risks.	☐ Mechanisms (e.g., training, documentation) exist for sustained and up-to-date awareness of safety and security guidance for all AI system stakeholders.

5.2.3 System risk assessments	☐ Agency personnel often use AI systems without adequate guidance on potential security issues, leading to additional risk exposure. ☐ The agency does not regularly identify or document the other IT systems within the agency that depend on or interface with the agency's AI systems.	☐ Some agency AI system operators are provided guidance on managing safety and security-related risks. ☐ The agency is in the process of identifying the other IT systems within the agency that depend on or interface with its AI systems. Efforts to do so are uneven across the agency.	☐ Risk controls for AI systems are consistently documented, disseminated, and stored in an accessible, centralized location to facilitate access. ☐ For the vast majority of AI systems, the other IT systems that are dependent on them or with which they interact are mapped, and the agency regularly uses these mappings to evaluate and prioritize risk mitigation activities.	☐ The risk management process proactively solicits and incorporates the views and experience of stakeholders, working with them to clarify use contexts, determine risks, identify potential benefits, and explore alternatives.

PAGE INTENTIONALLY BLANK

CNA

Dedicated to the Safety and Security of the Nation

www.cna.org

About CNA

CNA is a not-for-profit analytical organization dedicated to the safety and security of the nation. With nearly 700 scientists, analysts, and professional staff across the world, CNA's mission is to provide data-driven, innovative solutions to our nation's toughest problems. It operates the Center for Naval Analyses—the federally funded research and development center (FFRDC) of the Department of the Navy—as well as the Institute for Public Research. The Center for Naval Analyses provides objective analytics to inform decisionmaking by military leaders and ultimately improve the lethality and effectiveness of the joint force. The Institute for Public Research leverages data analytics and sophisticated methods to support federal, state, and local government officials as they work to protect the homeland, the American people, and industry.