DoD Artificial Intelligence Cybersecurity Risk Management Tailoring Guide

14 July 2025

Version 2

Executive Summary: Over the last few years, the DoD has prioritized digital modernization and adoption of artificial intelligence (AI) through various high-profile efforts. Throughout this period there has been a need to manage cybersecurity risk in AI systems. Consistent with Deputy Secretary of Defense direction via policy memorandums, DoD Instruction (DoDI) 8510.01 policy requirements, and integration of cybersecurity activities in the Adaptive Acquisition Pathways, this cybersecurity risk management tailoring guidance identifies the cybersecurity risk management activities, tools, teams, and processes that cybersecurity professionals need to integrate in the AI lifecycle. The content in this document is tailoring guidance and best practices. Policy requirements are cited where appropriate. DoD Components may implement cybersecurity risk management requirements in a manner they choose consistent with DoDI 8500.01, DoDI 8510.01, and Executive Order 13800.

As in the normal system development lifecycle, cybersecurity professionals need to be integrated as early as possible, so each lifecycle phase appropriately considers cybersecurity risks and mitigations. This in turn will allow for the best system posture, including informed test and evaluation (T&E), and support for an affirmative system cybersecurity assessment and authorization determination. Failure to appropriately integrate the following use case information and cybersecurity practices will jeopardize an AI systems' mitigation against cybersecurity risks and could impact operational use of AI systems.

Because AI system missions will vary, mission and system owners need to establish security objectives as early as possible. Cybersecurity professionals and even wider AI teams should reference Section 3, *Security Requirements for AI Systems, and Appendix B, System Security Requirements Mapping Tables,* as they progress through the AI lifecycle to ensure appropriate cybersecurity considerations are being applied to the AI system. While Section 3 describes the system risk management processes throughout the AI system lifecycle, Appendix B contains tables and lists outlining security priorities for cybersecurity professionals and data scientists or data engineers to consider when creating an AI system (i.e., infrastructure layer and AI model). Users should use this tailoring guide to accompany the Chief Digital and AI Officer *Responsible AI Toolkit* and the *DoD Strategy and Implementation Plan for Information and Communications Technology and Services Supply Chain Risk Management (ICT-SCRM) Assurance*.

Table of Contents

1. Introduction

This guidance, in support of DoD Instruction (DoDI) 8510.01, *Risk Management Framework (RMF) for DoD Systems*, establishes the DoD cybersecurity risk management tailoring guidance for the acquisition, development, use, sustainment, monitoring, and disposal of artificial intelligence (AI) systems as defined in Committee on National Security Systems (CNSS) Instructions 4009, *Committee on National Security Systems Glossary* (See Appendix C, *Glossary*). System owners should use this tailoring guidance to plan for and tailor the control mitigations related to the cybersecurity of AI systems. This also addresses the National Institute of Standards and Technology (NIST) AI RMF 1.0 – NIST AI 100-1 – direction for readers to consult the NIST Cybersecurity Framework and NIST Special Publication (SP) 800-37, Revision 2, to ensure AI systems are secure and resilient.

The DoD Chief Information Office (CIO) – in collaboration with the Offices of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) and Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) – has already published guidance integrating cybersecurity risk management, cyber test and evaluation (T&E), and acquisitions processes for the Software Acquisition Pathway as found on the DoD CIO Library in the section titled *Cybersecurity in the Adaptive Acquisition Framework*. AI system owners, data stewards, and data scientists or data engineers should leverage this published guidance to manage risks appropriate to the data and system's classification level – consistent with Executive Orders 13526 and 13556 – while adopting emerging DoD guidance related to test, evaluation, validation, and verification (TEVV) of models or Machine Learning Operations. This guidance seeks to help DoD organizations manage cybersecurity risks in the use of AI systems throughout the system lifecycle and thus encourage warfighter trust.

As this field continues to evolve, DoD CIO will partner with key stakeholders from the other Principal Staff Assistants to iterate upon this guidance. This tailoring guidance and the list of requirements in Appendix B are things that can be applied to AI. It is up to stakeholders to select which ones to apply when tailoring by deciding if they are appropriate and feasible to ensure the security at the classification level the AI system is intended to operate. This does not eliminate the requirement for all DoD systems, including AI systems, to be assessed and authorized.

1.1 Scope

This guidance applies to any AI system used or operated by DoD Components and presents tailored guidance for system owners and authorizing officials to use when authorizing an AI system for operational use. Figure 1 outlines how AI systems fit within the Department's Cybersecurity Program and the focused tailoring considerations needed for AI systems.

This guidance complements the existing DoD procedures for cybersecurity programs described in DoDI 8500.01, *Cybersecurity*, and DoDI 8510.01. This tailoring guidance identifies the cybersecurity activities that are most critical for meeting risk-based security. Consistent with the RMF process, this guidance helps system owners effectively manage security and privacy risks in diverse environments with complex and sophisticated threats, evolving missions and business functions, and changing system and organizational vulnerabilities.

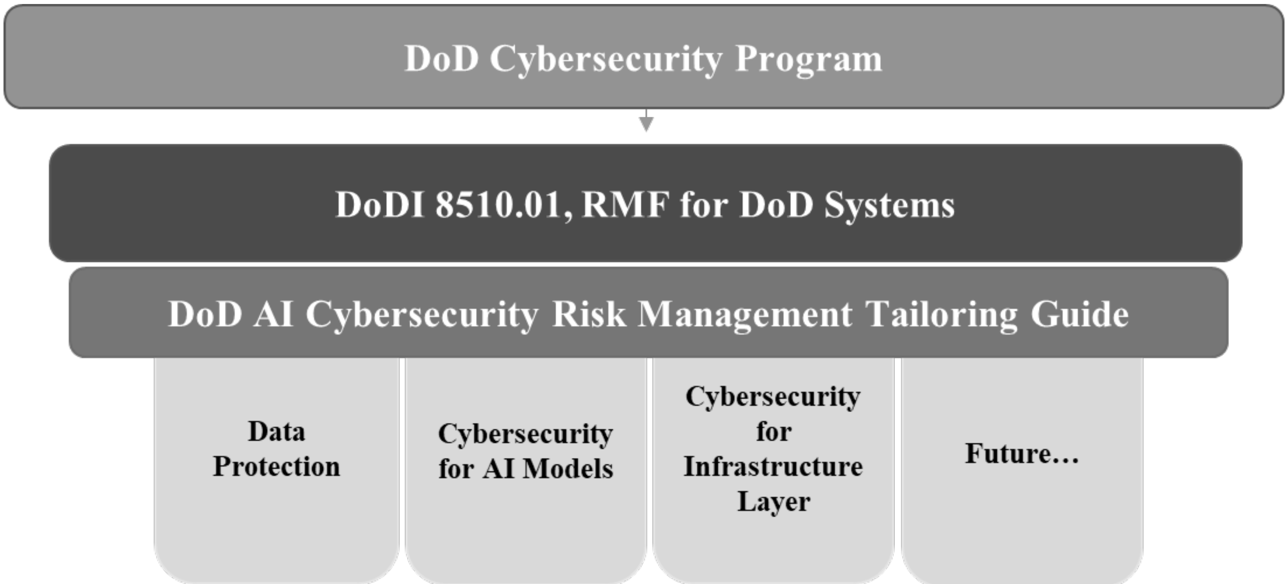

Figure 1, DoD Cybersecurity Program and Tailoring Considerations for AI Systems

DoDI 8500.01 requires DoD organizations to categorize all DoD systems in accordance with Committee on National Security Systems Instruction (CNSSI) 1253, *Categorization and Control Selection for National Security Systems*. DoD Memorandum, *Adoption of National Institute of Standards and Technology (NIST) SP 800-53 and CNSSI 1253 Revision 5*, 16 October 2023 requires DoD organizations to select and implement controls and control enhancements as published in CNSSI 1253, Revision 5, regardless of whether they are National Security Systems or not. This, and other cybersecurity implementation guidance can be found on the DoD Cybersecurity Knowledge Service (KS) (https://cybersecurityks.osd.mil/).

Each DoD organization retains the autonomy to determine its own risk tolerance for use of AI systems consistent with the requirements articulated by the *DoD Data, Analytics, and AI Adoption Strategy*, the *Responsible AI Strategy and Implementation Pathway*, Level II mission area owner risk guidance, the DoD 8500 policy issuance series, implementation guidance found on the DoD Cybersecurity KS, and the parameters of organization-specific cybersecurity programs. DoD organizations can adjust this tailoring guidance as needed to best support the needs of specific mission and business functions.

This document does not establish AI system performance expectations which are addressed in further detail in the Chief Digital and AI Officer (CDAO) *Responsible AI Toolkit* (https://rai.tradewindai.com/) and DoD Component specific AI use cases. Nor does this guide establish Zero Trust (ZT) guidance for AI systems. Implementing ZT will help secure DoD data, whether AI is involved or not. Thus, how ZT integrates with AI systems is not addressed here.

Designed to be general guidance for practitioners, programs, and organizations to implement for their specific AI systems, this guidance does not delve into classification differences other than to state that AI systems used in missions with a Sensitive Compartmented Information classification must follow existing DoD and Intelligence Community policies, as applicable.

1.2 Applicability

Consistent with DoDI 8510.01 applicability, the security priorities described in this guide apply to all AI systems operated by DoD or on behalf of the DoD by a contractor or other entity. This guidance does not apply retroactively to already-operational systems; however, DoD organizations should leverage this guidance as AI systems undergo updates, upgrades, and enhancements, where feasible. However, consistent with DoDI 8510.01 policy, organizations should apply this guidance as part of their annual control assessment review.

The traditional RMF (as described in NIST SP 800-37, Revision 2) and Assess Only processes (as described on the DoD Cybersecurity KS) apply to AI systems. However, AI systems have essential priorities and unique security considerations that require tailoring of the general DoD cybersecurity risk management methods. As related AI guidance is published by OUSD(R&E), OUSD(A&S), and the Office of the CDAO, this guidance for the cybersecurity of AI will need to stay synchronized. This aligns with the two Deputy Secretary of Defense signed policy memorandums – one establishing Task Force Lima and the other clarifying CDAO's role – affirming the distinct responsibilities of these DoD organizations in their Principal Staff Assistant roles (i.e., Deputy Secretary of Defense Memorandum, *Role Clarity for the Chief Digital and Artificial Intelligence Officer*, and Deputy Secretary of Defense Memorandum, *Establishment of Chief Digital and Artificial Intelligence Officer Generative Artificial Intelligence and Large Language Models Task Force, Task Force Lima*; see references).

This guidance establishes the DoD cybersecurity risk management tailoring guidance – including security and privacy controls – for the acquisition, development, use, sustainment, monitoring, and disposal of AI systems and increases users' ability to implement this risk management.

As such, this guidance describes security objectives tailored to the unique requirements of the continuous AI system lifecycle. Much like the Department's advancements in DevSecOps and Software Acquisition Pathway, which includes AI acquisition activities consistent with Secretary of Defense Memorandum, *Directing Modern Software Acquisition to Maximize Lethality*, the AI system lifecycle is not necessarily linear, and AI systems may change on an ongoing basis while in operations. For example, changes might occur due to periodic or continuous updates to the training data, addition of new AI models, or changes in technical approach.

This guidance supports DoD Component heads in their responsibility to provide protections, consistent with 44 U.S. Code Sec. 3554(a)(1)(ii) Federal Information Security

Management Act, for systems used or operated by an agency, contractor of an agency, or other organization on behalf of an agency.

1.3 Terms and Concepts

For purposes of this guidance, the term “AI system” is defined in alignment with CNSSI 4009. This definition incorporates the concepts of both the “system” definition from CNSSI 4009 and the NIST “system component” definition used in the RMF Process (See Appendix C, *Glossary*).

This guidance echoes terminology and concepts unique to specific AI systems while also relying on terms used throughout the cyber domain to better orient cybersecurity practitioners to the security needs of AI systems. This guide’s use of the term “organization” applies to any DoD organization that own and maintain responsibility for the cybersecurity of specific AI systems.

Appendix C, *Glossary*, contains definitions of essential characteristics of an AI system as found in CDAO publications, Executive Orders, or other DoD or U.S. Government issuances. DoD Components need to understand these considerations when integrating AI systems into operations to ensure systems’ functionality and security.

2. System Security Objectives for AI Systems

Security objectives consider the potential impact on confidentiality, integrity, and availability of information within a system as described in 44 U.S. Code, Sec. 3552*, Definitions*, and Federal Information Processing Standards Publication 199, *Standards for Security Categorization of Federal Information and Information Systems*.

To determine appropriate security objectives (i.e., categorization) for AI systems, data scientists or data engineers, acquisition personnel, and cybersecurity personnel – including cybersecurity engineers – need to identify the mission capabilities and development approach used. Just like any other system, to establish security objectives for AI systems, system owners and information owners need to emphasize completing RMF Prepare Step and Categorize Step tasks, especially Task P-12, *Information Types*, and P-13, *Information Life Cycle*, as early as possible (See the DoD Cybersecurity KS and NIST SP 800-37, Revision 2, for more information). The outcomes of these tasks will feed the subsequent RMF Steps throughout the system lifecycle.

Considering AI models require data security in all lifecycle stages, DoD organizations must protect the integrity and confidentiality of AI systems and the input, training, and output data feeding these systems. Failure to adequately address these security priorities may result in problems with AI models, including training models to make misinformed choices, providing biased outputs, or even allowing unauthorized personnel to view the model and its decision making. Though confidentiality and integrity are often the primary security objectives, AI systems also have availability requirements because AI systems can provide the warfighter with timely information in operations. Exact categorization for the different security objectives (Confidentiality-Integrity-Availability) will depend on AI systems' mission function, mission impact, and information processed, stored, or transmitted.

In alignment with DoDI 8510.01 policy and DoD Cybersecurity KS implementation guidance, it is recommended that as AI systems progress through the system development and AI lifecycles, system owners and cybersecurity teams evaluate the information outputs of these systems. This evaluation aims to determine whether the system should add different data types or be re-categorized at a different impact level. Such reviews should take place at least during every system re-authorization, if not sooner, as part of ongoing control assessments. The decision to re-categorize the AI system is subject to the discretion of the cognizant authorizing official. (More detailed information on AI system monitoring can be found in Section 3.1.6).

In addition to ensuring appropriate system confidentiality, integrity, and availability objectives, DoD organizations are releasing updated guidance and tools to manage the impact of AI systems on DoD operations. Users should use this document in tandem with other existing tools and policies, such as DoD Directive 3000.09, *Autonomy in Weapon Systems*, CDAO memorandum, *Interim Guidance on Use of Generative AI*, CDAO memorandum, *Guidelines and Guardrails to Inform Governance of Generative AI*, and CDAO's *Responsible AI Toolkit*. All of which specifically acknowledge the need to integrate with and adhere to cybersecurity policy and requirements.

3. Security Requirements for AI Systems

The following section takes a system-oriented approach to AI security. Rather than addressing security from an organizational perspective as in a Cybersecurity Framework Profile, this section describes implementation guidance for AI systems with relation to the controls established in CNSSI 1253. The following sections outline the security priorities data scientists or data engineers, and cybersecurity teams should consider when implementing cybersecurity for an AI system.

Using subject matter expert interviews within the Department, relevant sources (like the MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS) framework), and DoD CIO analysis, Appendix B contains tables mapping threats to AI system lifecycle phases and recommends security priorities (in terms of controls from CNSSI 1253) for organizations to consider when mitigating these threats. This guide heavily relies on ATLAS to address important parts of the total AI system attack surface. Ultimately, these security priorities will help organizations implement cybersecurity risk management for AI systems.

3.1 Cybersecurity Priorities in the AI System Lifecycle

At a minimum, risk management considerations for AI systems should include considerations for DoD systems and system components, using policy found in DoDI 8510.01 and guidance found on the DoD Cybersecurity KS. The following paragraphs outline the general cybersecurity priorities data scientists or data engineers, and cybersecurity teams should consider when creating and utilizing AI systems. System owners should address these system lifecycle concerns, consistent with DoDI 8580.01, and apply new AI system lifecycle guidance – consistent with Principal Staff Assistant authorities – from CDAO, OUSD(R&E), OUSD(A&S), and DoD CIO as it is published.

Additionally, organizational needs may require tailoring the security and privacy control baseline applicable to the AI system. However, this tailoring does not give organizations the freedom to accept unmanageable risks or skip steps in the RMF process. Instead, tailoring allows system owners to document deviations within the system's Security Plan thus allowing the authorizing official to make risk-informed decisions based on RMF results, T&E results, systems' unique mission/business functions, and the actions being performed by the AI system. If such deviations create unmitigated cybersecurity risks, those must be tracked and closed via the system's Plan of Action and Milestones (POA&M).

Because AI systems contain numerous parts operating as a whole AI system, control inheritance deserves attention from the entire cybersecurity team and AI model developers. For some AI system use cases, the system infrastructure layer may provide inheritance to the model. This allows data scientists or data engineers, and cybersecurity personnel involved in the model development to see vulnerabilities identified in the infrastructure layer, include appropriate mitigations, understand the system's risk posture, and actual inheritance the model can receive from the infrastructure layer.

3.1.1 Design and Develop AI Systems (i.e., Infrastructure Layer, Algorithms, Models, Data)

Just like in the DoD's guidance for integrating cybersecurity into Software Acquisitions, cybersecurity professionals – conducting Prepare and Categorize Step activities – should be engaged as early as possible in Design and Develop activities (see Figure 2). This ensures that the earliest design, acquisition, and custom development activities consider cybersecurity risks and priorities established in intake use cases and ideation mapping to existing systems. From a cybersecurity risk management perspective, there are unique threats around acquiring AI systems. These can include but are not limited to poisoned datasets used in model development, compromised hardware used as infrastructure for models and data, purchasing compromised commercial solutions, or utilizing vulnerable cloud or vendor architectures.

Involving cybersecurity professionals in setting requirements ensures cybersecurity is baked into Design and Develop activities. Requiring and communicating cybersecurity standards in contract language will ensure the Department can acquire the tools needed to enable AI system operations. Failure to implement these cybersecurity mechanisms could result in compromised datasets, backdoor exploits, malicious monitoring, model bias, or inefficient AI system operations.

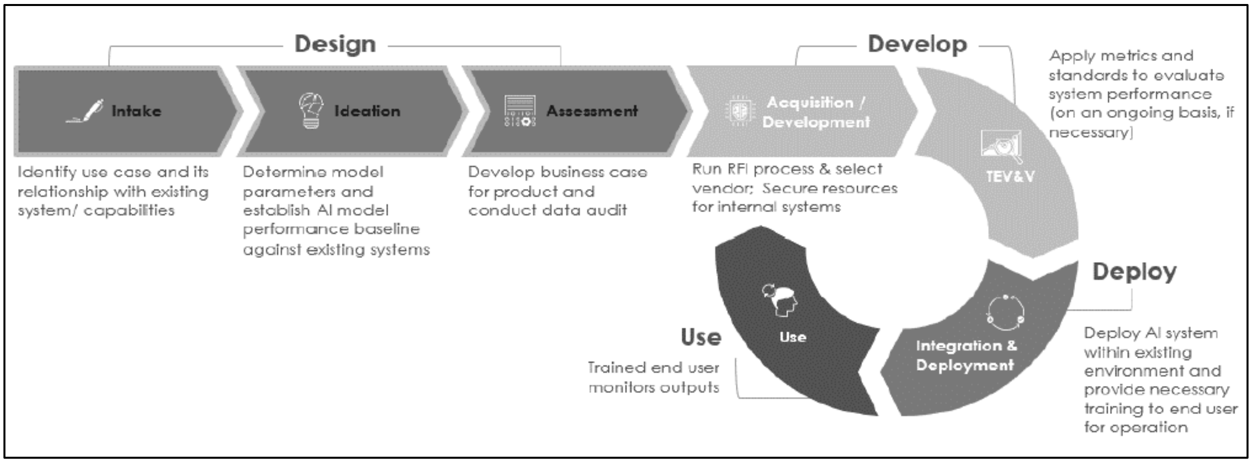

Figure 2: Responsible AI Activities throughout AI Lifecycle

Consistent with Executive Order 14306 and DODI 8531.01, *DoD Vulnerability Management*, successfully managing cybersecurity risk in AI systems requires the underlying infrastructure layer and AI models that enable them to be free of exploitable vulnerabilities and to function as expected and designed. Consistent with DoDI 5200.44 and DoDI 5000.83, DoD organizations must hold external and internal suppliers – in the DoD organization's supply chain – to the same security standards as that of the organization maintaining and using the AI system. This ensures DoD organizations manage technical risks from foreign intelligence collection, hardware and software vulnerabilities, supply chain exploitation, and reverse engineering of the components and systems that enable DoD warfighter capabilities and DoD-sponsored research.

System owners and AI teams should also consult DoD CIO policy and guidance on Information and Communications Technology-Supply Chain Risk Management (ICT-SCRM) to achieve confidence that products and services acquired and used in building DoD systems, networks, and applications are free of adversary influence at levels consistent with the sensitivity and criticality of the missions and functions performed or supported.

How AI systems (i.e., infrastructure layer and AI models) are acquired, and from whom, can lead to increased cybersecurity risk. The System and Services Acquisition (SA) and Supply Chain Risk Management (SR) control families contain controls intended to minimize these risks. These considerations should include data and its source. Documents outlining the origins and reliability of data will allow the organization to assess the attributes used in categorizing data and help in the categorizing the resulting AI system.

The AI team members listed in Section 3.1.3, *AI Model Development*, should begin developing data security plans for AI systems as soon as possible in the system design and planning stages before acquisition actions take place, conduct privacy assessments of the data used to train models, and identify any risks the data streams or development environment will pose to the completed AI model. Team members should treat data security plan creation as an ongoing process that should not be neglected and should evolve as an AI system's use case evolves. This does not mean teams can neglect establishing data security plans just because data security measures may change.

Although most of these threats can be mitigated by ensuring trust and integrity in the supply chain, testing of AI in an operationally representative contested or live, virtual, constructive environment is needed to ensure mitigation of the threats. As the DoD continues to advance adoption of AI, it should continue to refine standards around acquiring the technological elements of AI systems (i.e., infrastructure layer and AI models).

Instead of a single point in time, ICT-SCRM requires frequent review and due diligence upon releases of updated functionality. Additional ICT-SCRM guidance can be found in the recently published DoD CIO ICT-SCRM Strategy that aligns to Executive Order 14028. Additionally, CDAO has published the *Responsible AI Toolkit*, which includes considerations to include in acquisitions activities for program managers.

Appendix B, Tables 1-1 and 1-2 contain the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when acquiring AI systems.

3.1.2 Infrastructure Layer for AI System Development

Consistent with DoDI 8510.01, DoD organizations also need an authorized Design and Develop infrastructure layer for algorithms training to become AI models (see Figure 2) and have controls for input and movement between the environments with manual and automated code reviews.

Threats against Design and Develop infrastructure layers include (see Appendix B for more information):

- unauthorized access,
- injection attacks,
- data access attacks,
- evasion attacks, and
- attacks that could infer training data membership.

Since protecting data used in model training is a paramount security concern, authentication, provenance, configuration, physical, and audit controls protect information from unauthorized disclosure, access, or modification.

Organizations should limit system access to authorized users, service accounts (processes acting on behalf of authorized or privileged users), or authorized devices (including other systems) and limit the types of transactions and functions that authorized users and systems are permitted to exercise. Use of service accounts must follow existing operation orders, memorandums, or agency-specific policy.

These access control (AC) mitigations complement identification and authentication (IA) mitigations to act as gatekeepers for who, how, and when AI models can be developed and trained. In the AI development lifecycle stage, data and algorithm storage and transit protection is key. Consistent with the access requirements, AI systems also require transmission and information at rest protection to guarantee the confidentiality and integrity of information used to train AI models.

Appendix B, Table 2-1 and 2-2 contain the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when establishing infrastructure used in developing AI systems.

3.1.3 AI Model Development

Consistent with DoDI 8510.01, AI models need to complete the Assess Only Construct's Assess and Incorporate process, as found on the DoD Cybersecurity KS, to ensure cybersecurity requirements are identified, tailored appropriately, and assessed or evaluated before use. Such evidence should also include a Software Bill of Material consistent with Office of Management and Budget (OMB) 22-18, *Enhancing the Security of the Software Supply Chain through Secure Software Development Practices*. Additionally, the algorithms and data need to be assessed against controls like software and model integrity checks (e.g., SI-7), code review (e.g., SA-11(1)), and SCRM (e.g., SR-8). Failure to appropriately apply cybersecurity assessment to algorithms or data used to train models may result in inadvertent exposure to adversary injections or backdoors. Other key elements in mitigating threats to AI model training include auditing (e.g., AU-6) and monitoring (e.g., SI-4) controls. Users should refer to the *Responsible AI Toolkit* for more information on how to develop models (See Figure 2).

Modern Software Practice

Model development will follow a DevSecOps process (depicted in Figure 3) – or other modern software practice – guidance, tools, and processes consistent with current DoD policy, to ensure the delivered product has passed required security and functional tests to reduce the possibility of introducing vulnerabilities to the AI system. One such approach is to utilize continuous iteration-continuous delivery pipelines where changes in a specific iteration are tested prior to release in the production environment. While not the only software delivery methodology, DevSecOps is the preferred method for software delivery in the DoD and there are DevSecOps Reference Designs and other guidance on the DoD CIO Library page under Modern Software Practices (https://dodcio.defense.gov/library/).

These DevSecOps resources align with NIST SP 800-218, *Secure Software Development Framework Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities*, and Executive Order 14028, *Improving the Nation's Cybersecurity*. Users should also consult NIST SP 800-218A, *Secure Software Development Practices for Generative AI and Dual-use Foundation Models*, as appropriate.

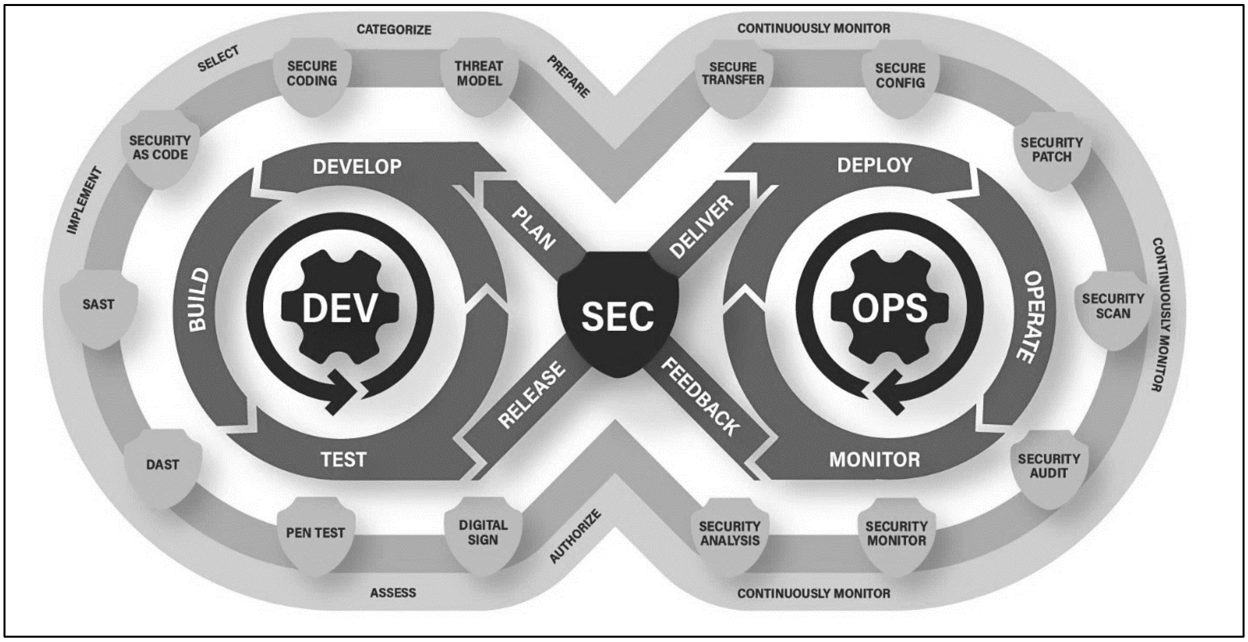

Figure 3, DevSecOps Distinct Lifecycle Phases and Philosophy

The team developing an AI model should make appropriate design considerations based on many factors, security being one of them. In one such possible situation, data scientists or data engineers may deploy algorithms and models to specific environments, whether in development or use, using containerization. In such use cases, organizations should adhere to the *Container Platform Security Requirements Guide*. Utilizing containers allows organizations to successfully roll back model functionality and mitigate against lost progress in model efficacy. However, teams are not limited to utilizing containers to deploy models especially as this field of technology is advancing at a rapid pace.

Models should be stored in a secure model catalog or repository, where they can be discovered and used on different missions sets or fine-tuned to new tasks.

Emphasis should be placed upon using authorized environments and valid risk management methods that address security concerns and meet requirements established in the RMF Prepare Step. Other potential use cases include air-gapping or network segmentation of an AI system training environment to address any concerns about negatively impacting high-performance computing resources.

AI System Team Members in Model Development

The scanning and hardening of models and the data involved is a collaborative process involving a host of special skillsets and expertise. After initial development and scans, AI teams should harden the model and remediate any findings. Consistent with DoDI 5000.89, *Test and Evaluation*, and the *Responsible AI Toolkit*, after hardening, the team (as outlined below) must perform rigorous data focused T&E to ensure the performance of the models is within accepted parameters, and ensure the model is performing with high efficacy. The following roles are essential to ensure data, model, and system security throughout the AI model development lifecycle (refer to the *Responsible AI Toolkit* for a list of personnel involved throughout the AI lifecycle):

1. Data scientists or data engineers: Data scientists and engineers acquire, prepare, and pre-process the data for AI model development. They play a crucial role in ensuring that sensitive or confidential information is properly handled and protected. They should also take steps to anonymize or de-identify data when necessary.

2. Data privacy officers or privacy teams: In organizations subject to data privacy regulations, data privacy officers and privacy teams are responsible for ensuring that AI projects adhere to DoD legal privacy and data provenance requirements. They assess the privacy impact of data usage and ensure that datasets are compliant with relevant DoD regulations.

3. Cybersecurity teams: Cybersecurity teams are responsible for assessing the security of data storage, model storage, data transmission, data access, and model access mechanisms. They play a role in implementing access controls, encryption, and other security measures to protect datasets from unauthorized access, breaches, and poisoning.

4. DoD project, data, and system owners: Project and data owners are typically individuals or departments responsible for the stewardship and governance of data and systems. They define access policies, grant permissions, and oversee the use of their data in AI training, validation, and testing. It is their responsibility to ensure that data is used in a secure and compliant manner.

5. AI teams: AI teams – including model designers and developers – should be aware of data security and model security best practices and ensure that these practices are followed during model development. They must also consider potential security risks associated with the model's output and predictions. See the *Responsible AI Toolkit* for more information on personnel involved in AI.

6. DoD legal teams: Legal teams can provide guidance on contracts, data sharing agreements, and liability issues related to data usage and AI model development. They ensure that legal agreements are in place to protect data and intellectual property rights.

7. End users and data subjects: Data security and model security also concerns end users and data subjects. End users and data subjects are responsible for abiding to law, regulation, policy, and Responsible AI practices when collecting and using data and models.

Appropriate personnel also need to develop verification and validation (V&V) and T&E plans in alignment with DoDD 3000.09 and DoDI 5000.89 (users should review DoDD 3000.09 for additional information on V&V). This V&V and T&E data should inform cybersecurity teams' execution of the Assess and Incorporate process as found on the DoD Cybersecurity KS. Further T&E guidance can be found in forthcoming T&E manuals published by OUSD(R&E).

Data Security

Data security plans should also consider risks of aggregating information consistent with DoD Manual 5200.01, Volume 3, *DoD Information Security Program: Protection of Classified Information,* DoD Manual 5205.02, *DoD Operations Security Program Manual,* and DoDI 8582.01, *Security of Non-DoD Information Systems Processing Unclassified Nonpublic DoD Information*. Such considerations should include discussions between cybersecurity professionals, data scientists or data engineers, relevant classification authorities, and data owners. This collaboration ensures those involved in AI system development, training, and use maintain appropriate information security requirements, user privileges, and data protection by sensitivity and classification level. Throughout the system lifecycle, it is essential for personnel to coordinate and maintain appropriate information security requirements, apply appropriate user privileges, and implement appropriate data protection requirements. System security assessments should include data security assessments, including risk assessments.

Security and privacy considerations should be integrated into the development process, and risk assessments should be conducted to identify and address potential vulnerabilities and threats related to the datasets and models used. Additionally, ongoing monitoring and auditing of data security practices are important to adapt to evolving risks and compliance requirements.

Cyber T&E and Cybersecurity Evidence

Each round of T&E includes applicable infrastructure layer and, if applicable, application scanning. If the scanning results in any high or critical findings, then system administrators need to continue hardening the infrastructure layer or application supporting model operations. After system administrators remediate findings, AI scientists or data engineers need to retest the model to ensure infrastructure layer or application changes do not adversely impact model operations. Consistent with DoDI 8510.01 policy, if unable to remediate these findings before deployment, authorizing officials can either not authorize the AI system or can justify the decision to deploy the AI system via POA&M documentation. This justification needs to note the potential risks involved with using the AI system with residual risks that cannot be remediated.

Leveraging the outcomes from RMF Process tasks P-7, Continuous Monitoring Strategy – Organization, and S-5, Continuous Monitoring Strategy – System, AI data scientists or data engineers and teams should conduct an iterative process of continuous monitoring, hardening, and testing to identify and remediate any risks or vulnerabilities. Initial validations include software integrity checks and vulnerability scans, which will identify any Common Vulnerabilities and Exposures (CVEs) or Control Correlation Identifiers, as appropriate. Specific T&E requirements and processes are covered by DoDI 5000.89 and appropriate T&E guidebooks and appendices. These scan, test, and validation results may reveal some known weaknesses or vulnerabilities that AI teams need to harden in the development infrastructure layer. This scanning is only an element in a comprehensive risk management process.

In accordance with DoDI 5000.83, OUSD(R&E) provides software and hardware assurance capabilities and expertise, including the Joint Engineering and Test Enterprise Portal's (JETEP) tool catalog offering a comprehensive list of security, assurance, protection, and testing tools available to the DoD community (https://jetep.apps.dso.mil/). These capabilities can support DoD AI programs to identify and mitigate vulnerabilities. Also, the JETEP has a body of knowledge, best practices, and guidance on AI assurance, to include runtime assurance for AI systems. In this way, system owners and data scientists or data engineers need to use AI scanning tools available through the JETEP. As of this moment, these AI specific tools do not replace the endpoint scanning tools used on the infrastructure layer of AI systems.

The CDAO team also uses a secure code scanning tool to conduct software assurance risk management, another tool as a vulnerability scanner in the development environment, and endpoint vulnerability scans in the production environment. Security Technical Implementation Guides (STIGs) and tools are environment based; organizations should apply STIGs as appropriate.

Once training is finished, AI data scientists or data engineers and the responsible cybersecurity team should include the model TEVV results in an AI system's security authorization package. Further information on model development can be found in the *Responsible AI Toolkit*.

The DoD CIO *DevSecOps Playbook* and DoD CIO Library guidance on integrating software acquisition activities with RMF processes provide helpful information to help teams validate the cybersecurity assessment, scanning, and T&E the model underwent.

Appendix B, Table 2-1 and 2-2 contains the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when using the DevSecOps or another modern software development process, consistent with current DoD CIO policy, to train and develop AI models.

3.1.4 AI System Deploy and Use

Deployed operational AI systems will include an AI model as well as the infrastructure layer that hosts the model and acts like a security wrapper by providing specific functionality requirements like performance (e.g., compute power) and security monitoring (e.g., protect, detect, respond, and recover from cybersecurity incidents) (see Figure 4).

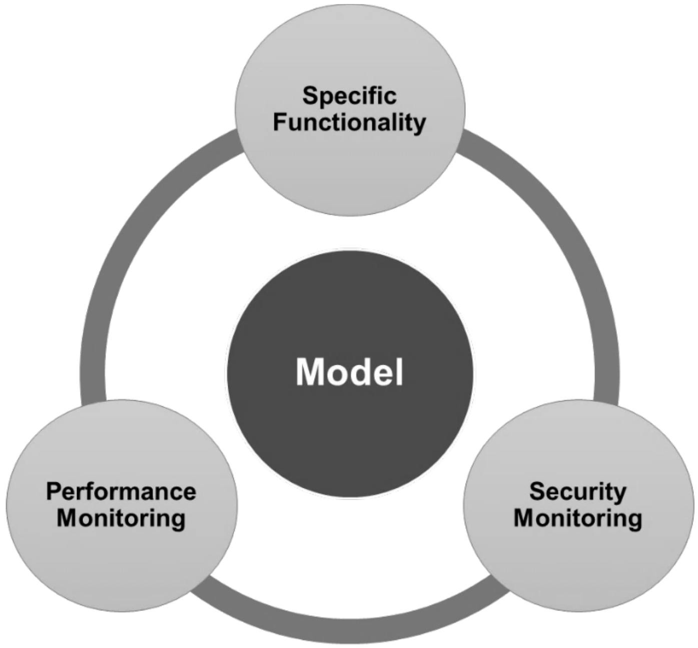

Figure 4, Functions of AI System Infrastructure Layer

Organizations may have different infrastructure layer use cases in operations and sustainment. Existing policy and guidance address the required authorization approach for each use case. Consistent with DoDI 8510.01, the following use cases all require the infrastructure layer hosting the model to have a system authorization that will include the model's (i.e., technology below the system level) cybersecurity assessment. A model's cybersecurity assessment evidence (i.e., Assess Only results and, when applicable, SBOM) is added to appropriate security authorization packages regardless of which infrastructure layer the model operates in. If a model is deployed to a development environment, see Section 3.1.3 for how to generate a body of evidence in a continuous manner.

Additionally, organizations should apply applicable STIGs and Security Requirement Guides to their AI systems and environments prior to deployment. For example, CDAO uses a tool to enable rapid transition and integration of AI models into operational environments. This tool adheres to the following STIGs:

- Application Security & Development
- Application Services
- High Availability Proxy
- Relevant operating system STIG or best practice (this should be used when checks are modified, or new checks are published)

Use Cases

Though not an exhaustive list for varying missions, different AI system use cases in operations and sustainment include:

- Hardware (e.g., on premises server):
    - System owners and teams need to follow the traditional RMF Process to authorize the hardware infrastructure layer for the AI system.
    - System owners and relevant teams need to conduct the Assess Only process for models utilizing best practices found on the DoD Cybersecurity Knowledge Service, in the *DevSecOps Playbook* and in the *Responsible AI Toolkit*. This evidence is added to the infrastructure layer's security authorization package.
    - The DoD Cybersecurity KS provides implementation guidance for the DoD RMF Process and Assess Only Construct.

- Cloud computing:
    - System owners and teams will need to leverage the latest version of the *Cloud Computing Security Requirements Guide* to authorize a cloud computing infrastructure layer.
    - The system owners and teams need to conduct the Assess Only process for models utilizing the best practices found on the DoD Cybersecurity KS, in *DevSecOps Playbook*, and in the *Responsible AI Toolkit*. This evidence is added to the cloud infrastructure layer's security authorization package.
    - The DoD Cybersecurity KS provides additional implementation guidance for cloud computing risk management.

- Hybrid cloud computing:
    - Authorizing officials are responsible for generating appropriate security authorization packages, including authorization determinations, for hybrid cloud environments.
    - To utilize a Hybrid Cloud Service Offering (CSO), an authorizing official will need an Enterprise to Public authorization, particularly for public CSOs. Mission Owners originate Enterprise to Public requests.
    - A CSO with a Federal Risk and Authorization Management Program (FedRAMP) authorization equates to a DoD Impact Level 2 (IL2) (i.e., public information) authorization through reciprocity.
    - A CSO adopting a FedRAMP authorization for DoD missions at a non-public level (i.e., Controlled Unclassified Information (CUI) or non-CUI) must also adopt FedRAMP+ controls specifically tailored to DoD. This would equate to a DoD Impact Level 4 (IL4). See the DoD Cloud Computing Cloud Service Provider SRG and DoD Cloud Computing Mission Owner SRG for more information on higher impact levels and FedRAMP+ controls.
    - The mission owner's authorizing official could leverage the DISA issued provisional authorization (PA) to issue a mission owner's authorization to operate (ATO) for hybrid use within their organization. DISA issued PAs can

be built on top of any FedRAMP authorization with additional FedRAMP+ controls.
   - System owners and relevant teams need to conduct the Assess Only process for models utilizing the best practices found on the DoD Cybersecurity KS, in *DevSecOps Playbook* and in the *Responsible AI Toolkit*. This evidence is added to the infrastructure layer's security authorization package.

- Weapons systems (including autonomous vehicles):
   - System owners and teams need to follow the traditional RMF Process to authorize weapon systems utilizing AI. System owners and teams should also refer to most updated version of the DoD Control Systems Security Requirements Guide for tailored cybersecurity risk priorities and mission objectives applicable to weapon systems, which are considered DoD control systems.
   - DoD personnel also need to adhere to DoD policy in DoDI 3000.09, *Autonomy in Weapon Systems*.
   - System owners and teams need to conduct the Assess Only process for models utilizing the best practices found on the DoD Cybersecurity KS, in *DevSecOps Playbook*, and in the *Responsible AI Toolkit*. This evidence is added to the infrastructure layer's security authorization package.

- Edge computing:
   - System owners and teams need to follow the existing DoDI 8510.01 policy on systems and technology below the system level.
   - Consistent with Assess Only process guidance found on the DoD Cybersecurity KS, if an edge device was assessed and approved for use via the Assess Only Construct, then the security authorization package containing the edge device's cybersecurity assessment should also include the model's cybersecurity assessment evidence.
   - If the edge device has its own security authorization package as a system, then the model cybersecurity assessment evidence should be added to that package.
   - Cybersecurity risk considerations should follow the cybersecurity risk management authorization decisions for a wholistic examination of cybersecurity risks to mission or business functions.

Integrating AI models into an operational status may include utilizing the Application Security & Development, Application Services STIGs as a best practice when checks are modified, or new checks are published. Deploying AI systems in research, engineering, prototyping, initial operational capability, or full operational capability use cases will also likely require organizations to implement strong configuration management (CM) security controls and permission settings through AC security controls.

Consistent with DoDI 8510.01 policy, all systems must receive a valid authorization before beginning operations. Systems that have skipped RMF Steps or do not have an adequate body of evidence via artifacts to support authorization determinations must capture deficiencies in a POA&M and be subjected to limited authorizations via an Authorization to Operate with Conditions.

Threats to AI systems (i.e., infrastructure layer and AI models) in operations and sustainment include data poisoning, inference attacks, model discovery, reverse engineering, and adversarial data manipulation. Appendix B, Table 3-1 and 3-2 contain the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when AI systems are operational or being sustained in a deployed status.

In addition to security considerations, organizations must also ensure they assess and appropriately address any privacy considerations for the AI system.

3.1.5 AI Model Deploy and Use

AI models face threats from model bias, degrade, drift, data poisoning, library vulnerabilities, and configuration error in operations and sustainment. This section addresses how to ensure AI system cybersecurity risk management in operations and sustainment.

Since models support different missions and use cases, and are trained on changing datasets, conditions for retraining models differ; however, defining these conditions – prior to operations and sustainment – is key to ensuring effective and reliable AI system operations. See the *Responsible AI Toolkit* for more information on the Responsible, Accountable, Supporting, Consulted, and Informed responsibility assignments for model retraining.

Per DoDI 8510.01, organizations must identify cybersecurity requirements, appropriately tailor controls, and assess the model's readiness for use in an operational environment. As an outcome, adding models to an already approved system will typically not require a new system authorization; however, consistent with DoDI 8510.01, cybersecurity teams must perform due diligence – consistent with the DoD Cybersecurity KS's Assess Only guidance – to ensure the model will not introduce unacceptable cybersecurity risk to system operations. The relevant authorizing official has final determination over the need for a new authorization decision. If there is a change in risk posture, the system should need a new authorization. Additionally, the model's acquisition source, training background, scan results, and cybersecurity T&E results should be added as evidence to a system's security authorization package. The DoD *DevSecOps Playbook* and DoD Enterprise DevSecOps Fundamentals explains how DevSecOps facilitates rapid and secure coding. Organizations should follow this guidance as closely as possible for models. Refer to Section 4 for additional information around system authorization considerations, including the need to apply the Assess Only Construct to models.

The point at which an organization chooses to stop models' learning has an impact on models' threat vectors and attack surface. Organizations must ensure they adhere to policy established in DoD Directive 3000.09 for autonomy in weapon systems. This decision on when learning stops is an operational risk consideration. Regardless of this decision, both configuration management and monitoring (e.g., SI-4) controls will be important to mitigate any threats to models' use in AI systems.

This guide brings attention to the need for AI system owners to discuss model training needs with qualified subject matter experts, and to be aware that those decisions may pose different cybersecurity risk management mitigations.

Appendix B, Table 3-1 and 3-2 contains the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when AI models, as part of AI systems, are operational or being sustained in a deployed status. Other mitigations against threats in the Use lifecycle phase include human responsibility for the AI systems' use as described in the DoD's Ethical Principles for AI.

3.1.6 AI System Monitoring

Relevant AI teams and cybersecurity professionals should develop and implement a suitable cybersecurity monitoring strategy (Task P-7, Continuous Monitoring Strategy – Organization, and Task S-5, Continuous Monitoring Strategy – System). The appropriate monitoring will vary based on the mission or business context, but this monitoring is required per DoDI 8510.01 and DoD Manual 8530.01. In certain use cases, authorizing officials may accept the risks of not monitoring AI systems, but such risk acceptance and reasoning must be documented and justified in a POA&M and Exception to Policy. Other mission or business context may necessitate code checks on a more frequent basis accounting for the wide variety of timescales on which evolution may be appropriate to mitigate cybersecurity risks to models in AI systems. Vulnerability and secure code scans should also occur whenever adding a new model to an AI system.

Of utmost importance to AI systems' use is the ability to monitor their performance and continuing security. Monitoring detects deviations from expected, trained behavior; potential spamming of the AI system with chaff data to influence outputs; and adversarial queries, data inputs, or other actions that can increase the cost of monitoring the AI system thus weaponizing the very function meant to detect malicious activity.

Consistent with OMB M-21-31, OMB A-130, DoD Manual 8530.01, and DoDI 8530.01, and DoDI 8510.01, once deployed, AI systems undergo Assured Compliance Assessment Solution scans, or the current endpoint security monitoring solution in use, per U.S. Cyber Command task orders and the responsible cybersecurity team must conduct, at least, annual security control reviews. AI system owners should apply system patches to address applicable Information Assurance Vulnerability Alerts for commercial off-the-shelf vulnerabilities and information assurance vulnerabilities, as disseminated by U.S. Cyber Command. As previously mentioned JETEP provides AI assurance tools also.

Additional steps identifying standards on monitoring frequency for performance and pre-established model efficacy thresholds will likely need to be established by CDAO and AI system owners. See the *Responsible AI Toolkit*.

Appendix B, Table 4-1 and 4-2 contains the cybersecurity priorities, in terms of CNSSI 1253 controls, for organizations to consider when monitoring AI systems.

3.1.7 AI System Decommissioning

AI systems require disposal and decommissioning consistent with the implementation guidance as found in the *Cloud Computing Security Requirements Guide* for cloud systems or the CNSSI 1253 for non-cloud systems. Additional decommissioning implementation guidance can be found on the DoD Cybersecurity KS and in the cybersecurity guidance for the Software Acquisition Pathway.

Because of the exposure to and large aggregation of data in these systems, proper sanitization and destruction is needed to ensure sensitive materials do not escape DoD control and become compromised by malicious actors or adversaries. Disposal activities need to account for the infrastructure layer as well as information related to the model, including the model's data, weights, T&E results, containers, and web applications used. Data at Rest destruction in the cloud is performed with the destruction of encryption keys. These disposal activities must also adhere to policy and procedures in National Security Agency/Central Security Service Policy Manual 9-12, *Storage Device Sanitization and Destruction Manual*.

4. Authorization Considerations

In the context of this guidance, a DoD organization's mission refers to the functions that organizations aim to accomplish via the use of AI systems. This may be the direct use of an AI system as the critical system for performing a business function or the use of an AI system as an element supporting the warfighting mission.

In addition to the existing cybersecurity risk management governance structures of the DoD Information Security Risk Management Committee (DoD ISRMC) and Defense Security/Cybersecurity Authorization Working Group (DSAWG), cybersecurity governance for AI systems also includes the organization which will use the AI system and the AI subject matter expertise assessment the model undergoes before being deployed.

Figure 5 portrays a notional example of how this authorization process takes place for the AI system's infrastructure layer and AI model. The infrastructure layer will follow the traditional RMF Process – including a mission owners expressed need for the AI system – while model development will occur in tandem utilizing the *Responsible AI Toolkit* and Assess Only Construct, as found on the DoD Cybersecurity Knowledge Service. Once model development is complete, the model's body of evidence – including its appropriate categorization recommendations – will be added to the infrastructure layer's security authorization package for review by the appropriate security control assessor before being sent to the authorizing official for a final determination.

To enable speed in AI system deployment, this guidance provides authorizing officials with a common understanding of the tools used in AI development (e.g., scanning, containers, DevSecOps) and a well-defined, understandable lexicon around AI development, use, and risk. This understanding and a well-developed, common lexicon will also help organizations establish a well-defined risk tolerance level for AI system operations. This guide prepares cybersecurity personnel and senior leaders to understand the unique security considerations and requirements needed for AI systems. An organization's culture towards adoption of automation and augmentation will also impact its ability to effectively deploy AI systems.

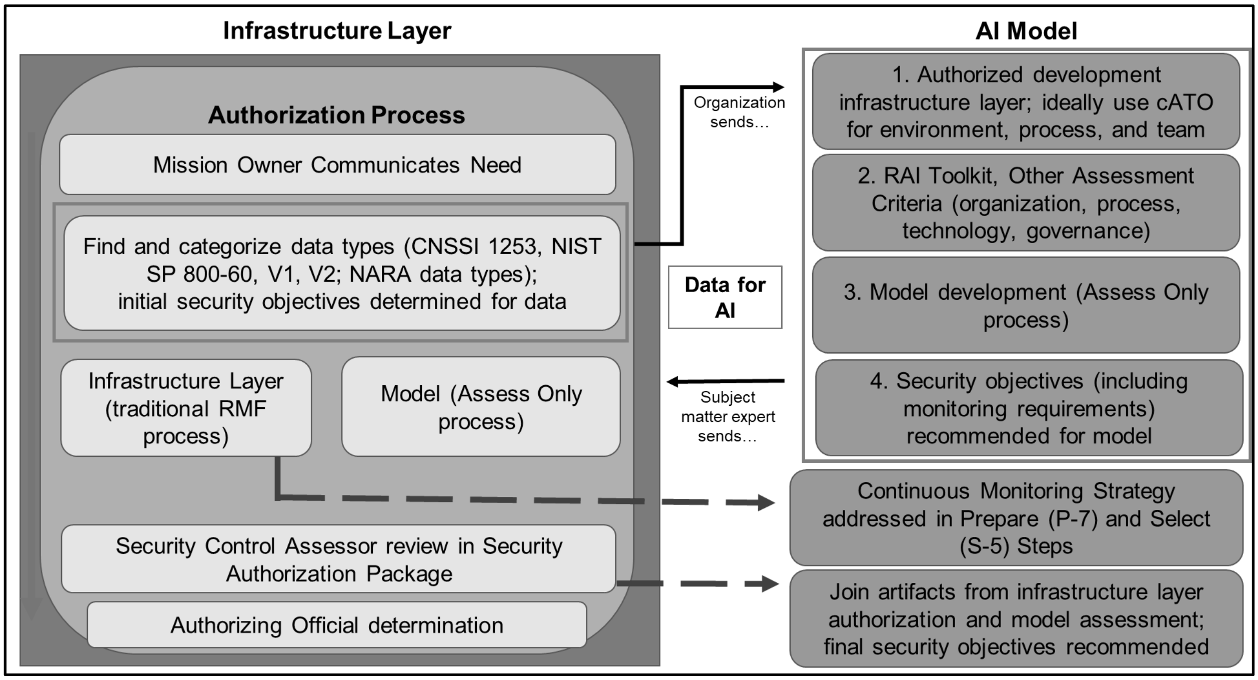

Figure 5, Notional AI System Authorization Process

Authorizing officials' risk tolerance is informed by things such as the maturity of the system, the mission and business functions performed by the system, and the system's – AI model and infrastructure layer – categorization. Ensuring appropriate CVEs are identified is another way to ensure authorizing officials have a full understanding of the risks and vulnerabilities in model use. Vulnerabilities may play a part in an authorizing officials' decision for risk tolerance, but they are only one component of a risk tolerance decision.

Authorizing officials are not responsible for model performance, but these metrics should inform risk decisions. Ensuring transparency in the model development and training allows authorizing officials to accept the risks of using AI systems because the evidence produced will allow them to understand its capabilities, risks, and hardening mitigations taken.

4.1 AI System Boundaries

All DoD systems, including AI systems, must have a valid authorization consistent with policy in DoDI 8510.01. Authorization is based on the boundary of the system. Organizations should also evaluate external connections to the authorization boundary (e.g., CA-3, *Information Exchange*) by assessing interconnections and dependencies of data streams that operationalize AI systems. Consider this exemplar:

1. AI Infrastructure Layer (at Boundary)
    - AI systems' infrastructure layer have an authorization boundary and require an authorization to operate (ATO).
    - Consistent with OMB A-130, DoDI 8500.01, DoDI 8530.01, and NIST SP 800-37, AI systems interacting with other systems external to the AI system's authorization boundary must have an interconnection security agreement in place detailing the system interaction and authorities.
    - This mitigates against organizational defenders mistaking AI system activities as malicious actors and ensures AI system activities are appropriately scoped.

2. AI Models (of a Larger System Boundary)
    - As part of an AI system, in a development, training, or operational status, an AI model – consisting of an algorithm and data set – leverage the Assess Only approach.
    - Consistent with DoDI 8510.01 and the implementation guidance found on the DoD Cybersecurity KS, the results of this cybersecurity assessment must be included in the final AI system security authorization package.

AI models do not need an ATO, but the actual system infrastructure layer does. Instead, AI models need cybersecurity evidence developed via the Assess Only Construct – this evidence should include a body of evidence supporting the cybersecurity assessment and should feature change management documentation, acquisitions documentation, and T&E results. Systems or cloud environments used to develop, deploy, and use AI models for use in DoD fall under DoDI 8510.01 policy.

Consistent with iterative development and security principles found in the *DevSecOps Playbook* and the DoD CIO's *Software Acquisition Pathway Integration with RMF* guidance, data scientists or data engineers and teams should work closely with the authorizing official to understand precisely what each control gate must validate before a model can be promoted to the next lifecycle phase. There is currently no one-size-fits-all answer to what cybersecurity criteria is sufficient, but cybersecurity assessment evidence – as developed via the DoD Cybersecurity KS Assess and Incorporate guidance for AI models – should support an authorizing official's reasonable acceptance of mitigation activities and residual risk, if any.

Per DoDI 8510.01, DoD organization can only operate authorized DoD systems with a current affirmative authorization decision – as issued by their Component's authorizing official – and need to maintain this authorization by operationalizing cybersecurity risk management. Mission risk will continue to be assessed and authorized by the authorizing official throughout the existence of an authorization. This is applicable at all system criticality levels.

As it continues to mature it's continuous monitoring of system risk, DoD seeks to enhance cybersecurity against expanding threats via a continuous ATO (cATO). This cATO effort is part of DoD CIO's and CDAO's cybersecurity for AI way ahead. Consistent with DoDI 8510.01 the DoD policy memorandum, *Continuous Authorization to Operate*, signed by the DoD Chief Information Security Officer, DoD CIO requires three main competencies that systems must possess to achieve a cATO – including active cyber defense and continuous monitoring.

DoD CIO has also released *cATO Evaluation Criteria for the DevSecOps Use Case* and a *DevSecOps Continuous Authorization Implementation Guide*. CDAO and DoD CIO will continue to work together to further define unique requirements and establish criteria for AI systems to achieve cATO.

4.2. Reciprocity for AI Systems

Establishing reciprocity for AI systems requires a review of the TEVV and assessment and authorization documentation – to include the appropriate body of evidence, security authorization package, system acquisition materials, and development processes and team in place – to ensure it meets the requirements and security objectives of the new use case. As with reciprocity for any system, failure to communicate the results, artifacts, and body of evidence generated from a system's authorization will hinder any sort of wide-scale rapid adoption of AI.

Consistent with DoD and CNSS policy, DoD organizations use reciprocity to reduce redundant testing, assessing, documenting, and the associated costs in time and resources. This is accomplished through sharing the system's body of evidence (e.g., RMF documentation) for authorizing officials' thoughtful, risk-based assessment on AI systems' applicability and suitability for a specific security landscape.

Users should refer to the *DoD Cybersecurity Reciprocity Playbook* on the DoD CIO Library and DoD Cybersecurity KS for more information on how to implement reciprocity.

Appendix A – References

44 U.S. Code 3554(a)(1)(ii), *Federal agency responsibilities*

44 U.S. Code 3552, *Definitions*

Office of Management and Budget M-22-18, *Enhancing the Security of the Software Supply Chain through Secure Software Development Practices*

Chief Digital and Artificial Intelligence Officer Memorandum, *Interim Guidance on Use of Generative AI*, 5 October 2023

Chief Digital and Artificial Officer Memorandum, *Guidelines and Guardrails to Inform Governance of Generative AI*, 12 July 2024

Chief Digital and Artificial Intelligence Officer, *Responsible Artificial Intelligence Toolkit*, 14 September 2023

Committee on National Security Systems Instruction 1253, *Categorization and Control Selection for National Security Systems*, 29 July 2022

Committee on National Security Systems Instruction 4009, *Committee on National Security Systems Glossary*, 2 March 2022

Deputy Secretary of Defense Memorandum, *Establishment of Chief Digital and Artificial Intelligence Officer Generative Artificial Intelligence and Large Language Models Task Force, Task Force Lima*, 10 August 2023

Deputy Secretary of Defense Memorandum, *Role Clarity for the Chief Digital and Artificial Intelligence Officer*, 1 February 2022

Secretary of Defense Memorandum, *Directing Modern Software Acquisition to Maximize Lethality*, 6 March 2025

DoD Chief Information Office Library, *DevSecOps Continuous Authorization Implementation Guide*, March 2024 <https://dodcio.defense.gov/library/>

DoD Chief Information Office Library, *DevSecOps Playbook*, as amended <https://dodcio.defense.gov/library/>

DoD Chief Information Office Library, *DoD cATO Evaluation Criteria: DevSecOps Use Case*, 29 May 2024 <https://dodcio.defense.gov/library/>

DoD Chief Information Office Library, *Software Acquisition Pathway Integration with Risk Management Framework*, 23 August 2023 <https://dodcio.defense.gov/library/>

DoD Chief Information Office Library, *DoD Cybersecurity Reciprocity Playbook*, March 2024 < https://dodcio.defense.gov/library/>

DoD Chief Information Office Library, *DoD Enterprise DevSecOps Fundamentals*, as amended <https://dodcio.defense.gov/library/>

DoD Strategy and Implementation Plan for ICT and Services Supply Chain Risk Management Assurance, June 2024 (https://cyber.mil/ict-scrm/resources-guidance/)

Intelink, *DoD Cyber Developmental Test and Evaluation Guidebook*, Version 3.0, June 2025 <https://go.intelink.gov/Cyber_DTE> (CAC-enabled)

DoD Cyber Exchange, *Application Security and Development Security Technical Implementation Guide*, as amended

DoD Cyber Exchange, *Container Platform Security Requirements Guide*, as amended

DoD Cyber Exchange, *DoD Cloud Computing Cloud Service Provider Security Requirements Guide*, as amended

DoD Cyber Exchange, *DoD Cloud Computing Mission Owner Security Requirements Guide*, as amended

DoD Cyber Exchange, *DoD Control Systems Security Requirements Guide*, as amended

DoD Directive 3000.09, *Autonomy in Weapon Systems*, 25 January 2023

DoD Directive 5200.47E, *Anti-Tamper (AT)*, Incorporating Change 3, 22 December 2020
DoD Instruction 5000.83, *Technology and Program Protection to Maintain Technological Advantage*, Change 1, 21 May 2021
DoD Instruction 5000.83, *Technology and Program Protection to Maintain Technological Advantage*, Incorporating Change 1, 21 May 2021
DoD Instruction 5000.89, *Test and Evaluation*, 19 November 2020
DoD Instruction 5200.39, *Critical Program Information (CPI) Identification and Protection Within Research, Development, Test, and Evaluation (RDT&E)*, Incorporating Change 3, 1 October 2020
DoD Instruction 5200.44, *Protection of Mission Critical Functions to Achieve Trusted Systems and Networks (TSN)*, 16 February 2024
DoD Manual 5200.01, Volume 3, *DoD Information Security Program: Protection of Classified Information*, Incorporating Change 3, 28 July 2020
DoD Manual 5205.02, *DoD Operations Security Program Manual*, Incorporating Chane 2, 29 October 2020
DoD Manual 8530.01, *Cybersecurity Activities Support Procedures*, 31 May 2023
DoD Instruction 8310.01, *Information Technology Standards in the DoD*, 7 April 2023
DoD Instruction 8330.01, *Interoperability of Information Technology, Including National Security Systems*, 27 September 2022
DoD Instruction 8500.01, *Cybersecurity*, Incorporating Change 1, Effective 7 October 2019
DoD Instruction 8510.01, *Risk Management Framework for DoD Systems*, 19 July 2022
DoD Instruction 8520.02, *Public Key Infrastructure and Public Key Enabling*, 18 May 2023
DoD Instruction 8530.01, *Cybersecurity Activities Support Procedures*, 31 May 2023
DoD Instruction 8531.01, *DoD Vulnerability Management*, 15 September 2020
DoD Instruction 8551.01, *Ports, Protocols, and Services Management*, 31 May 2023
DoD Instruction 8580.01, *Information Assurance (IA) in the Defense Acquisition System*, as amended
DoD Instruction 8582.01, *Security of Non-DoD Information Systems Processing Unclassified Nonpublic DoD Information*, 9 December 2019
DoD Memorandum, *Adoption of NIST Special Publication 800-53 and CNSSI 1253 Revision 5*, 16 October 2023
Executive Order 13526, *Classified National Security Information*, 29 December 2009
Executive Order 13556, *Controlled Unclassified Information*, 4 November 2010
Executive Order 13800, *Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure*, 16 May 2017
Executive Order 14028, *Improving the Nation's Cybersecurity*, 12 May 2021
Executive Order 14306, *Sustaining Select Efforts To Strengthen the Nation's Cybersecurity and Amending Executive Order 13694 and Executive Order 14144*, 6 June 2025
Federal Information Processing Standards Publication 199, *Standards for Security Categorization of Federal Information and Information Systems*, February 2004
MITRE Corporation, Adversarial Threat Landscape for Artificial-Intelligence Systems, as amended, <https://atlas.mitre.org/>
National Institute of Standards and Technology AI 100-1, *AI Risk Management Framework*, January 2023

National Institute of Standards and Technology Computer Security Resource Center Glossary, “system component,” <https://csrc.nist.rip/glossary/term/system_component>

National Institute of Standards and Technology Special Publication 800-37, Rev 2, *Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy*, December 2018

National Institute of Standards and Technology Special Publication 800-218, Version 1.1, *Secure Software Development Framework*, February 2022

National Institute of Standards and Technology Special Publication 800-218A, *Secure Software Development Practices for Generative AI and Dual-use Foundation Models*, July 2024

National Security Agency/Central Security Service Policy Manual 9-12, *Storage Device Sanitization and Destruction Manual*, 4 December 2020

Office of Management and Budget Circular A-130, *Managing Information as a Strategic Resource*, 28 July 2016

Office of Management and Budget Memorandum 21-31, *Improving the Federal Government’s Investigative and Remediation Capabilities Related to Cybersecurity Incidents*, 27 August 2021

Office of the Secretary of Dense Memorandum, *Continuous Authorization To Operate (cATO)*, 2 February 2022

Cybersecurity Knowledge Service, *Risk Management Framework (RMF) Assess and Incorporate*, as updated <https://cybersecurityks.osd.mil/dodcs/RMFImplementation/AssessOnly/Pages/AssessandIncorporate.aspx> (CAC-enabled)

Cybersecurity Knowledge Service, *Cloud Risk Management*, as updated <https://cybersecurityks.osd.mil/dodcs/rmffordodtech/pages/cloudriskmanagement.aspx> (CAC-enabled)

Cybersecurity Knowledge Service, *Common Security Controls and Inheritance*, as updated <https://cybersecurityks.osd.mil/dodcs/controlsandauthorization/securitycontrols/pages/commoncontrols.aspx> (CAC-enabled).

Cybersecurity Knowledge Service, *Categorize System - Overview*, as updated <https://cybersecurityks.osd.mil/dodcs/RMFImplementation/Categorize/Pages/CategorizeSystemOverview.aspx> (CAC-enabled)

Cybersecurity Knowledge Service, *RMF Process (DoD Systems)*, <https://cybersecurityks.osd.mil/dodcs/RMFImplementation/Pages/RMFProcess.aspx> (CAC-enabled)

Appendix B – System Security Requirements Mapping Tables

The following tables illustrate: 1) some possible threat vectors posed to AI systems and system components based on their lifecycle phases, 2) possible high-level security mitigations, as established by policy, that could help reduce cybersecurity risk to AI systems, and 3) specific security priorities for AI systems as represented by CNSSI 1253 controls. The mapping tables utilized threats found in the ATLAS framework and risk factors from the RMF Technical Advisory Group Secretariat's security analysis for software security, DevSecOps, and emerging capabilities like Robotic Process Automation. This appendix is not intended to be a wholistic, end all analysis of the threat area and available mitigations but acts as a reasonable starting point for authorizing officials and cybersecurity teams to consider in their cybersecurity risk management activities.

The threat vectors that follow are derived from ATLAS, which is a living knowledge base of adversary tactics and techniques against AI-enabled systems and is based on real-world attack observations and realistic demonstrations from AI red teams and security groups. For a more detailed description of threats, tactics, techniques, and procedures, users should go to the ATLAS website.

Table 1-1: Mapping AI Design and Develop Risks/Attack Vectors to Mitigations

Basic Threat Vector	***High-Level Mitigation***
1.1 Threat: Adversaries gain initial access to a system by compromising the unique portions of the supply chain. This can include hardware, data, the software stack, or the model itself. In some instances, the attacker will need secondary access to fully carry out an attack using compromised components of the supply chain.	Organizations must communicate standards and identify trustworthy sources and vendors for data, hardware, software stack, and algorithms utilized in AI systems consistent with DoDI 8310.01, DoDI 8500.01, DoDI 8510.01, DoDI 5000.83, DoDI 5200.39, DoDI 5200.44, and DoDI 5200.47E. An assessment from the intelligence community should also be performed to determine if this AI system provides the warfighter with a technical advantage. If so, protection methods should align with DoDI 5200.39 and DoDD 5200.47E.

Table 1-2: Security Priorities for AI Design and Develop

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
PM-9	Risk Management Strategy	X	X
PM-11	Mission and Business Process Definition	X	X
PT-1	Policy and Procedures		X
RA-3	Risk Assessment	X	
RA-3(1)	Risk Assessment | Supply Chain Risk Management		X
SA-4	Acquisition Process	X	X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SA-4(2)	Acquisition Process | Design and Implementation Information for Controls	X	X
SA-4(3)	Acquisition Process | Development Methods / Techniques / Practices	X	X
SR-2	Supply Chain Risk Management Plan	X	
SR-3	Supply Chain Controls and Processes		X
SR-4	Provenance		X
SR-4(3)	Validate as Genuine and Note Altered		X
SR-6	Supplier Assessments and Reviews	X	X
SR-6(1)	Supplier Assessments and Reviews | Testing and Analysis	X	
SR-9	Tamper Resistance and Detection		X

Table 2-1: Mapping AI Development Risks/Attack Vectors to Mitigations

Basic Threat Vectors	***High Level Mitigation***
2.1 Threat: **2.1.a. Model Poisoning:** Attacker gains access to training environment and adds data to original data set without altering original (Data Injection), modifies output labels and input data of original dataset (Data Manipulation), or alters the learning process or model itself (Logic Corruption). Modifying underlying data or its labels allows the adversary to embed vulnerabilities in ML models trained on the data that may not be easily detectable. The embedded vulnerability is activated later by data samples with an Insert Backdoor Trigger Poisoned data can be introduced via ML Supply Chain Compromise or the data may be poisoned after the adversary gains Initial Access to the system. **2.1.b. Unauthorized Access:** An unauthorized or malicious user accesses the training data or model (NIST definition) **2.1.c. Improper Configuration:** Incorrect system configuration allows for malicious or accidental alteration of data, algorithms, or models. **2.1.d. Data Access Attack:** Attacker gains access and uses training data to create a substitute model.	Organizations must secure the AI development and training environment with the appropriate configuration control, identification requirements, and cryptographic protections consistent with DoDI 8500.01 and DoDI 8530.01. When due diligence is done in procuring datasets and establishing training environments, model training must be monitored to ensure models are trained using the correct data and no system alterations allow for unwanted changes to the finished AI model.

Table 2-2: Security Priorities for AI Development

Control ID	**Control Name**	**Baseline**	
		Infrastructure Layer	**AI Models**
AC-1	Policy and Procedures	X	
AC-2	Account Management	X	X
AC-2(2)	Account Management | Automated Temporary and Emergency Account Management		X
AC-2(3)	Account Management | Disable Accounts		X
AC-2(4)	Account Management | Automated Audit Actions		X
AC-2(5)	Account Management | Inactivity Logout		X
AC-2(12)	Account Management | Account Monitoring for Atypical Usage		X
AC-3	Access Enforcement	X	X
AC-3(4)	Access Enforcement | Discretionary Access Control		X
AC-3(7)	Access Enforcement | Role-based Access Control	X	X
AC-4	Information Flow Enforcement	X	X
AC-5	Separation of Duties	X	X
AC-6	Least Privilege	X	X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
AC-6(8)	Least Privilege | Privilege Levels for Code Execution	X	X
AC-6(9)	Least Privilege | Log use of Privileged Functions		X
AC-6(10)	Least Privilege | Prohibit Non-Privileged Users from Executing Privileged Functions		X
AC-7	Unsuccessful Logon Attempts	X	X
AC-8	System Use Notification	X	X
AC-10	Concurrent Session Control	X	X
AC-12	Session Termination	X	X
AC-12(1)	Session Termination | User-initiated Logouts		X
AC-16	Security and Privacy Attributes	X	X
AC-17	Remote Access	X	X
AC-17(3)	Remote Access | Managed Access Control Points		X
AC-18	Wireless Access	X	
AC-20	Use of External Systems	X	
AC-21	Information Sharing	X	X
AC-23	Data Mining Protection	X	
AC-24	Access Control Decisions	X	X
AU-1	Policy and Procedures	X	X
AU-2	Events Logging	X	X
AU-3	Content of Audit Records	X	X
AU-3(1)	Content of Audit Records | Additional Audit Information		X
AU-4	Audit Log Storage Capacity		X
AU-4(1)	Audit Log Storage Capacity | Transfer to Alternate Storage		X
AU-5	Response to Audit Loging Process Failures		X
AU-5(1)	Response to Audit Loging Process Failures | Storage Capacity Warning		X
AU-5 (2)	Response to Audit Loging Process Failures | Real-Time Alerts		X
AU-6	Audit Review, Analysis, and Reporting	X	X
AU-6(4)	Audit Review, Analysis, and Reporting | Central Review and Analysis		X
AU-7	Audit Record Reduction and Report Generation	X	X
AU-8	Time Stamps	X	X
AU-9	Protection of Audit Information	X	X
AU-9 (3)	Protection of Audit Information | Cryptographic Protection		X
AU-10	Non-Repudiation	X	X
AU-12	Audit Record Generation	X	X
AU-12(3)	Audit Record Generation | Changes by Authorized Individuals		X
AU-13	Monitoring for Information Disclosure	X	X
AU-14	Session Audit	X	

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
AU-16	Cross-organizational Auditing Logging	X	
CA-2	Control Assessments	X	X
CA-8	Penetration Testing	X	X
CM-1	Policy and Procedures	X	X
CM-2	Baseline Configuration	X	X
CM-2(3)	Baseline Configuration | Retention of Previous Configurations	X	X
CM-2(6)	Baseline Configuration | Development and Test Environments	X	X
CM-3	Configuration Change Control	X	X
CM-3(2)	Configuration Change Control | Test / Validate / Document Changes	X	X
CM-3(7)	Configuration Change Control | Review System Changes	X	X
CM-4	Security Impact Analysis		X
CM-4(1)	Security Impact Analysis | Separate Test Environments	X	X
CM-4(2)	Security Impact Analysis | Verification of Security Functions	X	X
CM-5	Access Restrictions for Change	X	X
CM-5(1)	Access Restrictions for Change | Automated Access Enforcement and Audit Records		X
CM-5(4)	Access Restrictions for Change | Dual Authorization	X	X
CM-5(6)	Access Restrictions for Change | Limit Library Privileges	X	X
CM-7	Least Functionality	X	X
CM-7(2)	Least Functionality | Prevent Program Execution	X	X
CM-7(4)	Least Functionality | Unauthorized Software	X	X
CM-7(5)	Least Functionality | Authorized Software	X	X
CM-7(6)	Least Functionality | Confined Environments with Privileges	X	X
CM-7(7)	Least Functionality | Code Execution in Protected Environments	X	X
CM-7(8)	Least Functionality | Binary or Machine Executable Code	X	X
CM-8	System Component Inventory		X
CM-8(3)	System Component Inventory | Automated Unauthorized Component Detection		X
CM-9	Configuration Management Plan	X	X
CM-10	Software Usage Restrictions	X	X
CM-10(1)	Software Usage Restrictions | Open-source Software	X	X
CM-11	User-installed Software	X	X
CM-11(2)	User-installed Software | Software Installation with Privileged Status	X	X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
IA-3	Device Identification and Authentication	X	X
IA-3(1)	Device Identification and Authentication | Cryptographic Bidirectional Authentication		X
IA-7	Cryptographic Module Authentication	X	X
IA-9	Service Identification and Authentication	X	
IR-2	Incident Response Training		X
IR-3	Incident Response Testing		X
IR-4	Incident Handling		X
IR-4(6)	Incident Handling | Insider Threats	X	X
IR-5	Incident Monitoring	X	X
MA-3	Maintenance Tools	X	X
MA-4	Nonlocal Maintenance		X
MA-4(6)	Nonlocal Maintenance | Cryptographic Protection		X
MA-4 (7)	Nonlocal Maintenance | Disconnect Verification		X
PE-2	Physical Access Authorizations	X	
PE-3	Physical Access Control	X	
PE-6	Monitoring Physical Access	X	
PE-10	Emergency Shutoff	X	X
PL-2	System Security and Privacy Plans		X
PM-7	Enterprise Architecture	X	X
PS-3	Personnel Screening		X
PS-5	Personnel Transfer		X
PT-2	Authority to Process Personally Identifiable Information		X
PT-3	Personally Identifiable Information Process Purposes	X	X
PT-4	Consent		X
PT-5	Privacy Notice		X
PT-6	System of Records Notice		X
PT-7	Specific Categories of Personally Identifiable Information		X
PT-8	Computer Matching Requirements		X
RA-3	Risk Assessment	X	X
RA-3(1)	Risk Assessment | Supply Chain Risk Management		X
SA-4	Acquisition Process	X	X
SA-4(2)	Acquisition Process | Design and Implementation Information for Controls	X	X
SA-4(3)	Acquisition Process | Development Methods, Techniques and Practices	X	X
SA-8	Security and Privacy Engineering Principles		X
SA-8(1)	Security and Privacy Engineering Principles | Clear Abstraction		X
SA-8(2)	Security and Privacy Engineering Principles | Least Common Mechanism		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SA-8(3)	Security and Privacy Engineering Principles | Modularity and Layering		X
SA-8(4)	Security and Privacy Engineering Principles | Partially Ordered Dependencies		X
SA-8(5)	Security and Privacy Engineering Principles | Efficiently Mediated Access		X
SA-8(6)	Security and Privacy Engineering Principles | Minimized Sharing		X
SA-8(7)	Security and Privacy Engineering Principles | Reduced Complexity		X
SA-8(8)	Security and Privacy Engineering Principles | Secure Evolvability		X
SA-8(9)	Security and Privacy Engineering Principles | Trusted Components		X
SA-8(10)	Security and Privacy Engineering Principles | Hierarchical Trust		X
SA-8(11)	Security and Privacy Engineering Principles | Inverse Modification Threshold		X
SA-8(12)	Security and Privacy Engineering Principles | Hierarchical Protection		X
SA-8(13)	Security and Privacy Engineering Principles | Minimized Security Elements		X
SA-8(14)	Security and Privacy Engineering Principles | Least Privilege		X
SA-8(15)	Security and Privacy Engineering Principles | Predicate Permission		X
SA-8(16)	Security and Privacy Engineering Principles | Self-Reliant Trustworthiness		X
SA-8(17)	Security and Privacy Engineering Principles | Secure Distributed Composition		X
SA-8(18)	Security and Privacy Engineering Principles | Trusted Communications Channels		X
SA-8(19)	Security and Privacy Engineering Principles | Continuous Protection		X
SA-8(20)	Security and Privacy Engineering Principles | Secure Metadata Management		X
SA-8(21)	Security and Privacy Engineering Principles | Self-Analysis		X
SA-8(22)	Security and Privacy Engineering Principles | Accountability and Traceability		X
SA-8(23)	Security and Privacy Engineering Principles | Secure Defaults		X
SA-8(24)	Security and Privacy Engineering Principles | Secure Failure and Recovery		X
SA-8(25)	Security and Privacy Engineering Principles | Economic Security		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SA-8(26)	Security and Privacy Engineering Principles | Performance Security		X
SA-8(27)	Security and Privacy Engineering Principles | Human Factored Security		X
SA-8(28)	Security and Privacy Engineering Principles | Acceptable Security		X
SA-8(29)	Security and Privacy Engineering Principles | Repeatable and Documented Procedures		X
SA-8(30)	Security and Privacy Engineering Principles | Procedural Rigor		X
SA-8(31)	Security and Privacy Engineering Principles | Secure System Modification		X
SA-8(32)	Security and Privacy Engineering Principles | Sufficient Document		X
SA-8(33)	Security and Privacy Engineering Principles | Minimization		X
SA-9	External System Services		X
SA-11	Developer Testing and Evaluation	X	X
SA-11(1)	Developer Testing and Evaluation | Static Code Analysis	X	X
SA-11(2)	Developer Testing and Evaluation | Threat Modeling and Vulnerability Analysis		X
SA-22	Unsupported System Components		X
SC-2	Separation of System and User Functionality		X
SC-3	Security Function Isolation	X	X
SC-4	Information in Shared System Resources	X	X
SC-5	Denial of Service Protection		X
SC-5(1)	Denial-of-Service Protection | Restrict Ability to Attack Other Systems		X
SC-5(2)	Denial-of-Service Protection | Capacity, Bandwidth, and Redundancy		X
SC-5(3)	Denial-of-Service Protection | Detection and Monitoring		X
SC-7	Boundary Protection	X	X
SC-7(4)	Boundary Protection | External Telecommunications Services		X
SC-7(5)	Boundary Protection | Deny by Default – Allow by Exception		X
SC-8	Transmission Confidentiality and Integrity	X	X
SC-8(1)	Transmission Confidentiality and Integrity | Cryptographic Protection		X
SC-8(2)	Transmission Confidentiality and Integrity | Pre- and Post-Transmission Handling		X
SC-12	Cryptographic Key Establishment and Management	X	
SC-13	Cryptographic Protection	X	X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SC-18	Mobile Code		X
SC-18(1)	Mobile Code | Identify Unacceptable Code and Take Corrective Actions		X
SC-18(4)	Mobile Code | Prevent Automatic Execution		X
SC-23	Session Authenticity	X	X
SC-23(1)	Session Authenticity | Invalidate Session Identifiers at Logout		X
SC-23(3)	Session Authenticity | Unique System-Generated Session Identifiers		X
SC-23(5)	Session Authenticity | Allowed Certificate Authorities		X
SC-24	Fail In Known State		X
SC-28	Protection of Information at Rest	X	X
SC-28(1)	Protection of Information at Rest | Cryptographic Protection		X
SC-43	Usage Restrictions	X	
SI-2	Flaw Remediation	X	X
SI-2(5)	Flaw Remediation | Automatic Software and Firmware Updates		X
SI-2(6)	Flaw Remediation | Removal of Previous Version of Software and Firmware		X
SI-4	System Monitoring	X	X
SI-4(12)	System Monitoring | Automated Organization-Generated Alerts		X
SI-6	Security and Privacy Function Verification	X	X
SI-7	Software, Firmware, and Information Integrity	X	X
SI-7(1)	Software, Firmware, and Information Integrity | Integrity Checks		X
SI-7(2)	Software, Firmware, and Information Integrity | Automated Notifications of Integrity Violations	X	X
SI-7(3)	Software, Firmware, and Information Integrity | Centrally Managed Integrity Tools	X	X
SI-7(5)	Software, Firmware, and Information Integrity | Automated Response to Integrity Violations		X
SI-7(6)	Software, Firmware, and Information Integrity | Cryptographic Protection	X	
SI-7(8)	Software, Firmware, and Information Integrity | Auditing Capability for Significant Events		X
SI-7(12)	Software, Firmware, and Information Integrity | Integrity Verification	X	
SI-7(15)	Software, Firmware, and Information Integrity | Code Authentication		X
SI-12	Information Management and Retention		X
SI-12(1)	Information Management and Retention | Limit Personally Identifiable Information Elements		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SI-12(2)	Information Management and Retention | Minimize Personally Identifiable Information in Testing, Training and Research		X
SI-12(3)	Information Management and Retention | Information Disposal		X
SI-15	Information Output Filtering	X	X
SI-16	Memory Protection	X	X
SI-20	Tainting		X
SR-8	Notification Agreements	X	X

Table 3-1: Mapping AI System Deploy and Use Threat Vectors to Mitigations

Basic Threat Vector	*High Level Mitigation*
3.1 Threat: **3.1.a Inference Attacks: Model Inference API Access** **3.1.b Exfiltration via Inference API** **3.1.c Extract Model** **3.1.d Invert Model** **3.1.e Evasion Attacks: Evade Model** **3.1.f Denial of Service** **3.1.g Spamming of System with Chaff Data** **3.1.h Erode Model Integrity** **3.1.i Intellectual Property Theft** **3.1.j Cost Harvesting** **3.1.k Injection Attacks** **3.1.l Broken Authentication/Access Control** **3.1.m Misconfiguration** **3.1.n Continue Training After Deployment** **3.1.o ML-Enabled Product or Service** **3.1.p Physical Environment Access** **3.1.q Full Model Access** **3.1.r Discover Model Ontology** **3.1.s Discover Model Family** **3.1.t Train Proxy Model** **3.1.u Replicate Model** **3.1.v Verify Attack** **3.1.w. Infer Training Data Membership**	Organizations must secure the operations and sustainment environment of AI systems with the appropriate configuration control, identification requirements, cryptographic protections, contingency planning, scanning, and monitoring protections consistent with DoDI 8500.01, DoDI 8520.02, DoDI 8330.01, DoDI 8530.01, and DoD 8551.01. Adequately securing this operating space mitigates against threats posed by duplicating, degrading, or altering the AI system, which includes the AI system's infrastructure layer as well as the AI model.

Table 3-2: Security Priorities in AI Deploy and Use

Control ID	**Control Name**	**Baseline**	
		Infrastructure Layer	**AI Models**
AC-2	Account Management	X	X
AC-2(2)	Account Management | Automated Temporary and Emergency Account Management		X
AC-2(3)	Account Management | Disable Accounts		X
AC-2(4)	Account Management | Automated Audit Actions		X
AC-2(5)	Account Management | Inactivity Logout		X
AC-2(12)	Account Management | Account Monitoring for Atypical Usage		X
AC-3	Access Enforcement	X	X
AC-3(4)	Access Enforcement | Discretionary Access Control		X
AC-4	Information Flow Enforcement	X	X
AC-5	Separation of Duties	X	X
AC-6	Least Privilege	X	X
AC-6(9)	Least Privilege | Log use of Privileged Functions		X
AC-6(10)	Least Privilege | Prohibit Non-Privileged Users from Executing Privileged Functions		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
AC-7	Unsuccessful Logon Attempts	X	X
AC-8	System Use Notifications	X	X
AC-16	Security and Privacy Attributes	X	X
AC-17	Remote Access	X	X
AC-17(3)	Remote Access | Managed Access Control Points		X
AC-18	Wireless Access	X	
AC-20	Use of External Systems	X	
AT-2	Literacy Training and Awareness	X	X
AT-3	Role-Based Training	X	X
CA-2	Control Assessments	X	X
CA-3	Information Exchange	X	X
CA-7	Continuous Monitoring	X	X
CA-7(4)	Continuous Monitoring | Risk Monitoring		X
CM-2	Baseline Configuration	X	X
CM-2(3)	Baseline Configuration | Retention of Previous Configurations	X	X
CM-2(6)	Baseline Configuration | Development and Test Environments		X
CM-5	Access Restrictions for Change	X	X
CM-6	Configuration Settings	X	X
CM-7	Least Functionality	X	X
CM-7(1)	Least Functionality | Periodic Review	X	
CM-7(2)	Least Functionality | Prevent Program Execution		X
CM-7(4)	Least Functionality | Unauthorized Software	X	X
CM-7(5)	Least Functionality | Authorized Software	X	X
CM-8	System Component Inventory	X	X
CM-9	Configuration Management Plan	X	X
CM-11	User-Installed Software	X	X
CM-11(2)	User-Installed Software | Software Installation with Privileged Status	X	X
CP-2	Contingency Plan	X	
CP-3	Contingency Training	X	
CP-4	Contingency Plan Testing	X	
CP-9	System Backup	X	X
IA-2	Identification and Authentication (Organizational Users)	X	X
IA-2(1)	Identification and Authentication (Organizational Users) | Multifactor Authentication to Privileged Accounts		X
IA-2(2)	Identification and Authentication (Organizational Users) | Multifactor Authentication to Non- Privileged Accounts		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
IA-2(5)	Identification and Authentication (Organizational Users) | Individual Authentication with Group Authentication		X
IA-2(8)	Identification and Authentication (Organizational Users) | Access to Account-Replay Resistant		X
IA-2(12)	Identification and Authentication (Organizational Users) | Acceptance of PIV Credentials		X
IA-3	Device Identification and Authentication	X	X
IA-4	Identifier Management	X	
IA-5	Authenticator Management		X
IA-5(1)	Authenticator Management | Password-based Authentication		X
IA-5(2)	Authenticator Management | Public-Key based Authentication		X
IA-5(13)	Authenticator Management | Expiration of Cached Authenticators		X
IA-6	Authenticator Feedback		X
IA-7	Cryptographic Module Authentication	X	X
IA-8	Identification and Authentication (Non-Organizational Users)	X	X
IA-8(1)	Identification and Authentication (Non-Organizational Users) | Acceptance of PIV credentials from other agencies		X
IA-8(2)	Identification and Authentication (Non-Organizational Users) | Acceptance of External Party Credentials		X
IA-8(4)	Identification and Authentication (Non-Organizational Users) | Use of Defined Profiles		X
IA-9	Service Identification and Authentication	X	
IA-11	Reauthentication		X
IR-2	Incident Response Training		X
IR-3	Incident Response Testing		X
IR-4	Incident Handling	X	X
IR-4(2)	Incident Handling | Dynamic Reconfiguration	X	
IR-4(6)	Incident Handling | Insider Threats	X	X
IR-5	Incident Monitoring	X	X
IR-8	Incident Response Plan	X	
IR-9	Information Spillage Response	X	
IR-9(3)	Information Spillage Response | Post-spill Operations	X	
MP-2	Media Access	X	X
MP-4	Media Storage	X	X
MP-5	Media Transport	X	X
MP-6	Media Sanitization		X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
MP-7	Media Use	X	X
PE-2	Physical Access Authorization	X	
PE-10	Emergency Shutoff	X	X
RA-3	Risk Assessment	X	
RA-3(1)	Risk Assessment | Supply Chain Risk Management		X
RA-5	Vulnerability Monitoring and Scanning	X	X
SC-2	Separation of System and User Functionality		X
SC-3	Security Function Isolation	X	X
SC-4	Information in Shared Systems	X	X
SC-5	Denial-of-Service Protection	X	X
SC-5(1)	Denial-of-Service Protection | Restrict Ability to Attack Other Systems		X
SC-5(2)	Denial-of-Service Protection | Capacity, Bandwidth, and Redundancy		X
SC-5(3)	Denial-of-Service Protection | Detection and Monitoring		X
SC-7	Boundary Protection	X	X
SC-7(4)	Boundary Protection | External Telecommunications Services		X
SC-7(5)	Boundary Protection | Deny by Default – Allow by Exception	X	X
SC-13	Cryptographic Protection	X	X
SC-18	Mobile Code		X
SC-18(1)	Mobile Code | Identify Unacceptable Code and Take Corrective Actions		X
SC-18(4)	Mobile Code | Prevent Automatic Execution		X
SC-23	Session Authenticity	X	X
SC-23(1)	Session Authenticity | Invalidate Session Identifiers at Logout		X
SC-23(3)	Session Authenticity |Unique System-Generated Session Identifiers		X
SC-23(5)	Session Authenticity | Allowed Certificate Authorities		X
SC-24	Fail In Known State		X
SC-28	Protection of Information at Rest	X	
SC-28(1)	Protection of Information at Rest | Cryptographic Protection		X
SC-41	Port and I/O Device Access	X	
SI-2	Flaw Remediation	X	X
SI-2(5)	Flaw Remediation | Automatic Software and Firmware Updates		X
SI-2(6)	Flaw Remediation | Removal of Previous Version of Software and Firmware		X
SI-3	Malicious Code Protection	X	X
SI-4	System Monitoring	X	X

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
SI-4(12)	System Monitoring | Automated Organization-Generated Alerts		X
SI-6	Security and Privacy Function Verification	X	X
SI-10	Information Input Validation	X	X
SI-10(3)	Information Input Validation | Predictable Behavior		X
SI-11	Error Handling	X	
SI-12	Information Management and Retention		X
SI-12(1)	Information Management and Retention | Limit Personally Identifiable Information Elements		X
SI-12(2)	Information Management and Retention | Minimize Personally Identifiable Information in Testing, Training and Research		X
SI-12(3)	Information Management and Retention | Information Disposal		X
SI-15	Information Output Filtering	X	X
SI-16	Memory Protection	X	X
SI-20	Tainting		X

Table 4-1: Mapping AI Monitoring Threat Vectors to Mitigations

Basic Threat Vector	*High Level Mitigation*
4.1 Threat: **4.1.a Evade ML Model** **4.1.b Spamming ML System with Chaff Data** **4.1.c Cost Harvesting**	Organizations must ensure their AI systems have strong monitoring capabilities, consistent with DoDI 8500.01 and DoDI 8510.01, and have means to defend against adversarial tactics meant to misdirect monitoring.

Table 4-2: Security Priorities for AI Monitoring

Control ID	Control Name	Baseline	
		Infrastructure Layer	AI Models
CA-7	Continuous Monitoring	X	X
CA-7(3)	Continuous Monitoring | Trend Analyses	X	
CA-7(4)	Continuous Monitoring | Risk Monitoring		X
IR-5	Incident Monitoring		X
RA-5	Vulnerability Monitoring and Scanning	X	X

Appendix C – Glossary

Algorithm. A method or set of rules or instruction to be followed in calculations or other problem-solving operations, particularly by a computer. (Source: DARPA/*DoD Responsible AI Strategy and Implementation Pathway*)

Artificial intelligence (AI). AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems. (Source: *DoD AI Strategy*)

Artificial Intelligence (AI) system. An AI system is a computer software running on physical hardware, or other context, that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical networks. (Source: Adapted from CNSSI 4009)

Autonomy. Autonomy refers to a system's ability to accomplish goals independently, or with minimal supervision from human operators in environments that are complex and unpredictable. (Source: DARPA/DoD Responsible AI Strategy and Implementation Pathway)

Availability. Ensuring timely and reliable access to and use of information. (Source: CNSSI 4009)

Data card. A document for a dataset that provides insight into collection, processing, usage, and security practices. (Source: *DoD Responsible AI Strategy and Implementation Pathway*)

Data element. A basic unit of information that has a unique meaning and subcategories (data items) of distinct value. Examples of data elements include gender, race, and geographic location. (Source: NIST Glossary)

Information and communications technology (ICT) supply chain risk management (SCRM). Also called "cyber SCRM," it is the process of identifying, assessing, and mitigating the supply chain risks associated with the development and use of distributed and interconnected information technology/operational technology and ICT product and service supply chains. The term is derived and modified for DoD use from National Institute of Standards and Technology documents and publications, to include National Institute of Standards and Technology Special Publication 800-161 and National Institute of Standards and Technology Interagency or Internal Report 8276. (Source: DoDI 5200.44)

Infrastructure Layer. The hosting environment for the AI system, explicitly providing compute, storage, network resources, and additional managed services to enable functional, cybersecurity, and non-functional capabilities. (Derived from *DoD Enterprise DevSecOps Fundamentals*, Version 2.1, September 2021)

Interconnection security agreement (ISA). A security document that specifies the technical and security requirements for establishing, operating, and maintaining the interconnection. It also supports the memorandum of understanding/agreement between the organizations. Specifically, the ISA documents the requirements for connecting the IT systems, describes the security controls that will be used to protect the systems and data, contains a topological drawing of the interconnection, and provides a signature line. (Source: CNSSI 4009/NIST Special Publication 800-47)

Machine Learning (ML). The study or the application of computer algorithms that improve automatically through experience. ML algorithms build a model based on training data in order to perform a specific task, like aiding in prediction or decision-making processes, without necessarily being explicitly programmed to do so. (Source: National Security Commission on AI Final Report)

Supply chain. A system of organizations, people, activities, information, and resources, possibly international in scope, that provides products or services to consumers. (Source: CNSSI 4009).

System. Any organized assembly of resources and procedures united and regulated by interaction or interdependence to accomplish a set of specific functions (Source: CNSSI 4009).

System component. A discrete identifiable information technology asset that represents a building block of a system and may include hardware, software, and firmware. (Source: NIST Glossary)

Trust. Trust is established by ensuring that AI systems are cognizant of and are built to align with core values in society, and in ways which minimize harms to individuals, groups, communities, and societies at large. Defining trustworthiness in meaningful, actionable, and testable ways remains a work in progress. In part, we rely on the practice of trustworthy computing as adopted by some in computer science and system engineering fields—"trustworthiness of a computer system such that reliance can be justifiably placed on the service it delivers (IEEE)"; "of an item, ability to perform as and when required (ISO/IEC/IEEE)". On other hand, the AI user trust decision, as other human trust decisions, is a psychological process. There is currently no method to measure user trust in AI or measure what factors influence the users' trust decisions. (Source: *DoD Responsible AI Strategy and Implementation Pathway*)

Appendix D – Revision History

Version	Date	Page(s) Changed	Comments
1.0	02 July 2024	N/A	Original Baseline Document
2.0	14 July 2025	All	Incorporated minor administrative edits better aligning to OPM memorandums, Executive Orders, and DoD policy. The most significant change is the inclusion of controls to consider when assessing the cybersecurity of AI Models via the Assess Only Construct (see Appendix B).