

NIST SP 800-53 Control Overlay for Securing AI Systems: Using and Fine-Tuning Predictive AI

Annotated Outline for Cyber AI Profile Workshop #2

Note to Reviewers

To facilitate discussion and engage interested stakeholders at the [Cyber AI Profile Workshop #2](#) on January 14, 2026, NIST is issuing this annotated outline of proposed content for **NISTIR 8605, Control Overlays for Securing AI Systems: Overview and Methodology**, and **NISTIR 8605A, Control Overlays for Securing AI Systems: Using and Fine-Tuning Predictive AI**.

The authors welcome feedback at the workshop on January 14, through ongoing engagement in the Control Overlays for Securing AI Systems (COSAiS) Slack Channel, and by email to overlays-securing-ai@list.nist.gov by February 13, 2026. Submission of feedback by this date will allow it to be considered as the authors develop the initial public draft; additional opportunities to comment on subsequent drafts will be available to all stakeholders.

Feedback on any of the content and related topics in this annotated outline are welcomed by attending the Cyber AI Profile Workshop #2 working sessions, through the COSAiS Slack channel, and by email. NIST is specifically interested in feedback on the following:

- The proposed NISTIR volumes and development timeline.
- The proposed development methodology and iterative approach to developing the series of overlays.
- Development assumptions, specifically the assumption that the organization has implemented the NIST SP 800-53B moderate impact security control baseline. Alternative approaches to identifying controls that are “assumed” to be already implemented by the organization are encouraged.
- Updates to the proposed use cases for the *Using and Fine-Tuning Predictive AI Overlay*.
- Proposed structure and content, both technical and level of detail for the overlays.
- Specific feedback on the *Using and Fine-Tuning Predictive AI Overlay* initial content and additional selected controls for inclusion in the overlay.

The content of this discussion draft/annotated outline will be included into NISTIR 8605 and NISTIR 8605A.

Learn more about the Control Overlays for Securing AI Systems (COSAiS) Project at <https://csrc.nist.gov/Projects/cosais>.

Proposed Deliverables and Timeline

NIST plans to issue a series of NIST Interagency Reports (IR) to address control overlays for securing AI systems. The proposed publications and planned release order in this series are:

1. NISTIR 8605, Control Overlays for Securing AI Systems: Overview and Methodology
2. NISTIR 8605A, Control Overlays for Securing AI Systems: Using and Fine-Tuning Predictive AI
3. NISTIR 8605B, Control Overlays for Securing AI Systems: Adapting and Using Generative AI
4. NISTIR 8605C, Control Overlays for Securing AI Systems: Security Controls for AI Developers
5. NISTIR 8605D, Control Overlays for Securing AI Systems: Using Agentic AI: Single Agent and Multi-Agent

NIST intends to issue NISTIR 8605 and NISTIR 8605A as drafts for public comment by Q3 FY2026 and continue release additional volumes for drafts through 2026 and 2027, refining previous draft volumes based on ongoing lessons learned and feedback. The series (all volumes) will be finalized in 2027, pending available resources. The overlays (Volumes A-D) will be available in the publications and made available online in different data formats through the NIST Cybersecurity and Privacy Reference Tool when finalized.

Background

- **NISTIR 8605:** This section will provide background on why this series and overlays were created and an overview of the types of AI systems addressed in the multivolume series. This section will include a high-level overview of the NIST SP 800-53 controls, what overlays are and how they can be used.
- **NISTIR 8605A:** This section will focus on background on predictive AI systems, an overview of common attacks and mitigations for predictive AI systems.

Development Approach and Assumptions

- This section will be included in NISTIR 8605.
- Background on Overlays:
 - o Offers organizations additional customization options for control baselines and may be a fully specified set of controls, control enhancements, and other supporting information (e.g., parameter values) derived from the application of tailoring guidance to SP 800-53B control baselines or derived independently of control baselines.
 - o Provides an opportunity to build consensus across communities of interest and develop a starting point of controls that have broad-based support for very specific circumstances, situations, and/or conditions.
- The overlays in this series assume the implementing organization has a risk management process and cybersecurity program in place (i.e., it is assumed that a process to select, implement, assess, determine and manage cybersecurity risk, existing cybersecurity

controls are in place organization wide, such as policy and procedures, and controls to manage risk to other enterprise systems).

- It is assumed that the AI systems are categorized as ([FIPS 199](#)) moderate impact for security, and the [NIST SP 800-53B](#) moderate impact baseline is selected and implemented.¹
- For organizations utilizing different types of AI (predictive, generative, etc.), overlays can be used collectively to manage risks. The control sets for each overlay will be distinct for the application of the control to address specific risks associated with the different types of AI.
- These use cases will address cybersecurity risks at three stages of the AI life cycle: (i) model training, (ii) model deployment, and (iii) model maintenance. Each scenario will address a different business workflow scenario (e.g., a specific application of an on-premises or third-party hosted AI model, using propriety data or publicly available data).
- This overlay may include control enhancements without the base control selected. In this scenario, it is assumed the organization has implemented the base control as part of the underlying IT system supporting the machine learning functionalities.
- The overlays are developed by:
 - o Mapping [NIST AI 100-2e2025 AI Taxonomy](#) to [NIST SP 800-53 controls](#) for attacks/mitigations
 - o Identifying controls that are assumed to be implemented by organization (i.e., included in the NIST SP 800-53B Moderate Impact Security Control Baseline)
 - o Tailoring controls for the use case²

Overlay Characteristics

- This section will be included in NISTIR 8605A.
- This overlay identifies only the controls to address specific risks for usage of **Predictive AI**
- This overlay addresses the use of predictive AI to automate 2 different business workflows (use cases³):
 - A. (On-premises model and third-party (cloud) model, proprietary data): Support of Operations and Supply Chain – Predictive Maintenance. Use of predictive AI to analyze sensor or continuous monitoring data of systems/system components to predict failures, automate work orders for preventative maintenance. Update the model based on actual maintenance needed.
 - B. (On-premises model and third-party (cloud) model, public-facing input data): Candidate Recruitment Automation. Use of predictive AI to scan candidate resumes, predict candidate success, and automate initial screening to streamline the hiring process. Update the model after successful or unsuccessful interviews and hires.

¹ NIST intends to revisit this assumption and revise as needed based on workshop and written feedback.

² The selected controls are tailored for each use case; organizations consider the need to further refine based on their own implementation.

³ The use cases has been updated from the [Concept Paper](#) to use a two use cases for both on-premises and third-party models and propriety and public-facing input data instead of four.

Applicability

- This section will be included in NISTIR 8605A.
- The primary focus on this overlay is to provide additional implementation guidance on securing the use of predictive AI systems. The controls and control enhancements from this overlay are tailored and applied on an as-needed basis by the implementing organization.

Overlay Overview and How to Use

- This section will be included in both NIST 8605 and NISTIR 8605A.
- Placeholder for overview of overlays, additional information about overlay structure (see example below) and how to use.

Explanation of Key Terms

- Different stages of the AI lifecycle used in the Overlay Summary Table (model training, model deployment, model maintenance, continuous)
- On-premises model, third-party hosted model, proprietary information, public information

Proposed Example Structure for Overlay Controls⁴

Control ID, Control Name e.g., AC-06, Least Privilege
Selected in SP 800-53B Moderate Baseline: {{select one}: Yes; No}
Applicable AI Lifecycle Phase(s): {{select one or more}: Model Training; Model Deployment; Model Maintenance; Continuous}
Assumptions: Any relevant assumptions for selection and implementation, including the implementation of a base control (for control enhancements)
Control Tailoring: (Note that not all controls will include all potential tailoring considerations identified below) <ul style="list-style-type: none"> - Recommendations for organization-defined parameters - Modifications to the control language - Modifications to the Discussion to include relevant technical guidance on applicability or considerations for on-prem/third-party deployment, and proprietary or public information throughout the applicable AI lifecycle phases - Suggestions for additional related controls (not included in SP 800-53) - Additional references to assist with implementation
Relevant NIST AI 100-2e2025 Attack ID: e.g., NIST AML.011

How to Use

- Placeholder for examples and suggestions of how to use the overlay.

⁴ Note that this is the narrative structure proposed for the NISTIR document series to facilitate ease of review and providing feedback. When the NISTIR series is finalized, the overlays will be released in different data formats to facilitate ease of use.

Using and Fine-Tuning Predictive AI Systems Controls

- This section will be included in NISTIR 8605A.

Summary Table

Note to Reviewers
<i>All selected controls for the overlay will be listed in this summary table in the order they appear in SP 800-53. The table below is populated with only a subset of example controls for the annotated outline.</i>

Control / Control Enhancement	AI Lifecycle Phase	Overlay Tailoring		
		Control Requirement	Organization-Defined Parameter	Discussion
AC-06 , Least Privilege	<input type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input checked="" type="checkbox"/> Continuous			X
CM-02 , Baseline Configuration	<input checked="" type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input checked="" type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X
CM-04 , Impact Analyses	<input checked="" type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input checked="" type="checkbox"/> Model Maintenance <input checked="" type="checkbox"/> Continuous			X
RA-05 , Vulnerability Monitoring and Scanning	<input type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input checked="" type="checkbox"/> Continuous		X	X
SA-11(02) , Developer Testing and Evaluation Threat Modeling and Vulnerability Analyses	<input checked="" type="checkbox"/> Model Training <input checked="" type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X
SA-15(01) , System and Services Acquisition Quality Metrics	<input checked="" type="checkbox"/> Model Training <input checked="" type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X
SA-15(08) , System and Services Acquisition Reuse of Threat and Vulnerability Information	<input checked="" type="checkbox"/> Model Training <input checked="" type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X

DRAFT ANNOTATED OUTLINE – January 2026
NIST SP 800-53 Control Overlays for Securing AI Systems: Using Predictive AI

Control / Control Enhancement	AI Lifecycle Phase	Overlay Tailoring		
		Control Requirement	Organization-Defined Parameter	Discussion
SC-05(03) , Denial-of-Service Protection Detection and Monitoring	<input type="checkbox"/> Model Training <input checked="" type="checkbox"/> Model Deployment <input checked="" type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X
SC-07(10) , Boundary Protection Prevent Exfiltration	<input type="checkbox"/> Model Training <input checked="" type="checkbox"/> Model Deployment <input checked="" type="checkbox"/> Model Maintenance <input type="checkbox"/> Continuous			X
SI-03(08) , Malicious Code Protection Detect Unauthorized Commands	<input type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input checked="" type="checkbox"/> Continuous			X
SI-04(02) , System Monitoring Automated Tools and Mechanisms for Real-Time Analysis	<input type="checkbox"/> Model Training <input type="checkbox"/> Model Deployment <input type="checkbox"/> Model Maintenance <input checked="" type="checkbox"/> Continuous			X

Note to Reviewers

The list of controls identified for potential inclusion in this overlay is not complete. The intent of this annotated outline is to provide some illustrative examples of the level of detail proposed for the overlay.

NIST welcomes feedback on all aspects of the overlay and specifically seeks input on the structure of the overlay, proposed controls and control enhancements for inclusion, level of detail included in the overlay outline, and any potential gaps (to include the opportunity to add new controls/control enhancements not included in SP 800-53).

Additional controls and control enhancements proposed for inclusion in this overlay include:

- [AC-03, Access Enforcement](#)
- [AC-22, Publicly Accessible Content](#)
- [AU-02, Event Logging](#)
- [AU-06, Audit Record Review, Analysis, and Reporting](#)
- [CA-03, Information Exchange](#)
- [CA-07, Continuous Monitoring](#)
- [CA-08, Penetration Testing](#)
- [CM-03, Configuration Change Control](#)
- [CM-04\(01\), Impact Analyses | Separate Test Environments](#)
- [CM-05, Access Restrictions for Change](#)
- [CM-06, Configuration Settings](#)
- [PE-09\(02\), Power Equipment and Cabling | Automatic Voltage Controls](#)

DRAFT ANNOTATED OUTLINE – January 2026
 NIST SP 800-53 Control Overlays for Securing AI Systems: Using Predictive AI

- [PE-14\(02\), Environmental Controls | Monitoring with Alarms and Notifications](#)
- [PT-03\(01\), Personally Identifiable Information Processing Purposes | Data Tagging](#)
- [PT-04, Consent](#)
- [PT-06, System of Records Notice](#)
- [PT-07, Specific Categories of Personally Identifiable Information](#)
- [RA-03\(02\), Risk Assessment | Use of All-Source Intelligence](#)
- [SA-08\(14\), Security and Privacy Engineering Principles | Least Privilege](#)
- [SA-08\(19\), Security and Privacy Engineering Principles | Continuous Protection](#)
- [SA-08\(22\), Security and Privacy Engineering Principles | Accountability and Traceability](#)
- [SA-11\(05\), Developer Testing and Evaluation | Penetration Testing](#)
- [SA-11\(09\), Developer Testing and Evaluation | Interactive Application Security Testing](#)
- [SA-15\(02\), Development Process, Standards, and Tools | Security and Privacy Tracking Tools](#)
- [SA-15\(06\), Development Process, Standards, and Tools | Continuous Improvement](#)
- [SA-17\(03\), Developer Security and Privacy Architecture and Design | Formal Correspondence](#)
- [SC-05, Denial-of-Service Protection](#)
- [SC-23, Session Authenticity](#)
- [SC-28, Protection of Information at Rest](#)
- [SC-32, System Partitioning](#)
- [SC-39, Process Isolation](#)
- [SC-43, Usage Restrictions](#)
- [SI-03, Malicious Code Protection](#)
- [SI-04, System Monitoring](#)
- [SI-07, Software, Firmware, and Information Integrity](#)
- [SI-07\(05\), Software, Firmware, and Information Integrity | Automated Response to Integrity Violations](#)
- [SI-10, Information Input Validation](#)
- [SI-10\(03\), Information Input Validation | Predictable Behavior](#)
- [SI-10\(05\), Information Input Validation | Restrict Inputs to Trusted Sources and Approved Formats](#)
- [SI-12\(02\), Information Management and Retention | Minimize Personally Identifiable Information in Testing, Training and Research](#)
- [SI-15, Information Output Filtering](#)
- [SI-17, Fail-Safe Procedures](#)
- [SI-18, Personally Identifiable Information Quality Operations](#)
- [SI-19\(06\), De-Identification | Differential Privacy](#)
- [SI-19\(07\), De-Identification | Algorithms and Software](#)
- [SI-23, Information Fragmentation](#)
- [SR-04, Provenance](#)
- [SR-11, Component Authenticity](#)

Access Control

Control ID: AC-06, Least Privilege
Selected in SP 800-53B Moderate Baseline: Yes
Applicable AI Lifecycle Phase(s): Continuous
Assumptions: AI system access control may be integrated into the organization-wide (enterprise) ICAM implementation.
Control Tailoring: [Discussion] To ensure the model is only available to authorized and appropriate individuals, the organization identifies and authorizes the minimum set of access rights for users, services, models, and pipelines to perform the assigned functions of the predictive AI system. Least

privilege applies to humans, non-human identities to include machines (e.g., services, AI agents) and data access throughout the AI model lifecycle. Organizations consider assigning unique, non-shared identities to training pipelines, inference services, and model monitoring agents, and restricts service accounts to only specific datasets, model artifacts and compute resources. It is expected that the access and permissions required by humans and non-human identities acting on behalf of an individual will change throughout the lifecycle.

Relevant NIST AI 100-2e2025 Attack ID: NIST AML.011, NIST AML.013

Configuration Management

Control ID: [CM-02, Baseline Configuration](#)

Selected in SP 800-53B Moderate Baseline: Yes

Applicable AI Lifecycle Phase(s): Model Training; Model Maintenance

Assumptions: AI system configuration management may be integrated into organization-wide (enterprise).

Control Tailoring:

[Discussion] Predictive AI systems may introduce additional configuration elements that are not explicitly addressed in baseline configurations of enterprise IT systems to potentially include machine learning frameworks and libraries, model architectures, and certain compute environments. The organization considers AI components and dynamic models in scope for the baseline configuration, including infrastructure components, the AI software stack, data and pipeline configurations, and model configuration artifacts.

Relevant NIST AI 100-2e2025 Attack ID: NIST AML.024, NIST AML.026

Control ID: [CM-04, Impact Analyses](#)

Selected in SP 800-53B Moderate Baseline: Yes

Applicable AI Lifecycle Phase(s): Continuous

Assumptions: Impact analyses for the AI system can leverage and build on existing organization-wide or system processes.

Control Tailoring:

[Discussion] Minor configuration changes could significantly alter model behavior for predictive AI systems, potentially introducing bias, drift or impacting the accuracy of the outputs. The impact analysis is expanded to include analysis of the infrastructure and platform system components, software (to include machine learning frameworks, libraries, and containers), and data pipeline for security and privacy impact, model behavior and overall risk.

Relevant NIST AI 100-2e2025 Attack ID: NIST AML.022

Risk Assessment

Control ID: RA-05, Vulnerability Monitoring and Scanning
Selected in SP 800-53B Moderate Baseline: Yes
Applicable AI Lifecycle Phase(s): Continuous
Assumptions: Vulnerability Monitoring and Scanning can leverage and build upon existing organization-wide or system processes and tools.
<p>Control Tailoring: [Organization-Defined Parameters]</p> <ul style="list-style-type: none"> [Assignment: organization-defined frequency and/or randomly in accordance with organization-defined process]: at system deployment, after significant AI model or pipeline changes, at organization-defined intervals based on risk assessment <p>[Discussion] The potential attack surface for predictive AI systems differs from traditional, enterprise IT systems. Organizations consider how to include and tailor existing processes and tools for the unique components of predictive AI systems such as machine learning frameworks, specialized libraries that are updated frequently, dynamic compute resources, different model artifacts.</p>
Relevant NIST AI 100-2e2025 Attack ID: NISTAML.024, NISTAML.026

System and Services Acquisition

Control ID: SA-11(02), Developer Testing and Evaluation Threat Modeling and Vulnerability Analyses
Selected in SP 800-53B Moderate Baseline: No
Applicable AI Lifecycle Phase(s): Model Training; Model Deployment
Assumptions: The threat modeling and vulnerability analysis methods applied to the AI system are consistent whether the development and training of the AI system is performed within the system authorization boundary, performed by a separate development group within the organization, or is performed as part of a product delivered by a third-party vendor.
<p>Control Tailoring: [Discussion] From development to deployment, developers update their understanding of the threats that can affect the deployment and track the vulnerabilities associated with the model, data sets, and tools used to create the predictive AI system environment. Given the scope of Predictive AI systems, additional focus may be applied to the system services supporting the machine learning.</p>
Relevant NIST AI 100-2e2025 Attack ID: NISTAML.013, NISTAML.022, NISTAML.05, NISTAML.051

Control ID: SA-15(01), Development Process, Standards, and Tools Quality Metrics
Selected in SP 800-53B Moderate Baseline: No
Applicable AI Lifecycle Phase(s): Model Training; Model Deployment

<p>Assumptions: The quality metrics applied to the AI system are consistent whether the development and training of the AI system is performed within the system authorization boundary, performed by a separate development group within the organization, or is performed as part of a product delivered by a third-party vendor.</p>
<p>Control Tailoring: <i>[Discussion]</i> Metrics can include acceptable levels of trust and reliability associated with the training data and the machine learning algorithms as benchmarks for resilience against known attack types. Periodic reviews of the model against the metrics can establish ongoing efforts to build security into the model. Models are not expected to be deployed unless they meet the organization-defined quality metrics.</p>
<p>Relevant NIST AI 100-2e2025 Attack ID: NISTAML.022</p>

<p>Control ID: SA-15(08), Development Process, Standards, and Tools Reuse of Threat and Vulnerability Information</p>
<p>Selected in SP 800-53B Moderate Baseline: No</p>
<p>Applicable AI Lifecycle Phase(s): Model Training; Model Deployment; Model Maintenance</p>
<p>Assumptions: The use and reuse of threat and vulnerability information applied to the AI system are consistent whether the development and training of the AI system is performed within the system authorization boundary, performed by a separate development group within the organization, or is performed as part of a product delivered by a third-party vendor.</p>
<p>Control Tailoring: <i>[Discussion]</i> Lessons learned from existing predictive AI deployments can be included in the development toolset for future development and guide the establishment of standardized processes for the development of new predictive AI systems. Understanding the threats and vulnerabilities associated common to the type of system and the development tools enables developers to establish “best practices” for securing the training model and associated applications.</p>
<p>Relevant NIST AI 100-2e2025 Attack ID: NISTAML.022</p>

System and Communications Protection

<p>Control ID: SC-05(03), Denial-of-Service Protection Detection and Monitoring</p>
<p>Selected in SP 800-53B Moderate Baseline: No</p>
<p>Applicable AI Lifecycle Phase(s): Model Deployment; Model Maintenance</p>
<p>Assumptions: IT services provide denial-of-service protections to mitigate availability risks to the system.</p>
<p>Control Tailoring: <i>[Discussion]</i> Protect against data being introduced to the model that can result in corruption of the data, uncontrolled queries, or introduces addition perturbations of testing samples that can affect system performance relative to performance objectives and thresholds. Monitoring for activity outside of known parameters and performance thresholds can be an indicator of a</p>

potential attack that should result in notifications and incident response measures. A key objective is to ensure that the model is not resulting in a self-inflicted DoS resulting from uncontrolled iterations updating the training model.
Relevant NIST AI 100-2e2025 Attack ID: NISTAML.031

Control ID: SC-07(10), Boundary Protection Prevent Exfiltration
Selected in SP 800-53B Moderate Baseline: No
Applicable AI Lifecycle Phase(s): Model Deployment; Model Maintenance
Assumptions: The impact from data exfiltration may be lower during development and training activities where exfiltration may be preventable using IT methods. In the deployment phase, the AI system boundary protections are implemented to prevent the exfiltration of the deployed model algorithms, data structures, and training data as a function of application usage.
<p>Control Tailoring: <i>[Discussion]</i> The exfiltration of information about the training model can lead to the development of more sophisticated attacks that subvert the ability of the model to meet objectives and requirements. Additional protections are needed to prevent the exfiltration of information that can be used to engineer data privacy attacks and model privacy attacks.</p> <p>Tests for exfiltration can include methods to identify potential seeding by an attacker, identifying where an attacker is able to introduce data into the model that enables the extraction of additional information from the training model data sets. The frequency of testing may depend on the sensitivity of the data being used in the model and the criticality of the system relative to organization and system objectives for the system.</p> <p><i>“extraction attacks are more successful when the model is seeded with more specific and complete information — the more the attacker knows, the more they can extract”</i> (NIST AI 100-2e2025, 3.3.2)</p>
Relevant NIST AI 100-2e2025 Attack ID: NISTAML.03, NISTAML.031, NISTAML.032, NISTAML.033, NISTAML.034)

System and Information Integrity

Control ID: SI-03(08), Malicious Code Protection Detect Unauthorized Commands
Selected in SP 800-53B Moderate Baseline: No
Applicable AI Lifecycle Phase(s): Continuous
Assumptions: Malicious code protections may already exist within the organization and system environments to address specific IT-related threats.
<p>Control Tailoring: <i>[Discussion]</i> Detecting unauthorized activity for all access points of the model, whether as it is being trained or once it is deployed, it a critical element of preventing the release of information about the model that can be used to subvert to trustworthiness and reliability of the system objectives. Data ingest methods should be able to parse input to ensure that system level commands are not being submitted, much in the way that SQL injection attacks are identified.</p>

While submitted commands may not be able to be executed based on the submission path, the submission still represents an attempt to gain a response from the system that an attacker can use to refine their attack strategy. Understanding the relationship of the machine learning tools, the model, and the data storage components can enable the identification and disabling of malicious code before it can subvert any portion of the system.

Relevant NIST AI 100-2e2025 Attack ID: NISTAML.026

Control ID: [SI-04\(02\), System Monitoring | Automated Tools and Mechanisms for Real-Time Analysis](#)

Selected in SP 800-53B Moderate Baseline: Yes

Applicable AI Lifecycle Phase(s): Continuous

Assumptions: System monitoring tools are in place to analyze all IT-related activity.

Control Tailoring:

[Discussion] System monitoring tools track interactions with all data sets and algorithms that are applied to the development and training of the model, and to all activity associated with the model once deployed. Automated methods identify attempts to subvert the model (i.e., poisoning, evasion) and distinguishes activity outside of established parameters for the model responses which can indicate attempts by an attacker to extract information about the model which can be used to craft future attacks. Appropriate personnel are notified of alerts based on organization-defined prioritizations and thresholds associated with the severity of the identified alert.

Relevant NIST AI 100-2e2025 Attack ID: NISTAML.031

Terms and Definitions

Note to Reviewers

This section will include terms and deliverables specific to the NISTIR 8605 volume.

Additional Information or Instructions

Note to Reviewers

This section will include additional information or instructions specific to the NISTIR 8605 volume