Check for updates

**NIST Internal Report**
**NIST IR 8596 iprd**

# Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)

*NIST Community Profile*

Initial Preliminary Draft

Katerina Megas
Barbara Cuthill
Marissa Dotter
Michael Garris*
Ishika Khemani
Bronwyn Patrick
Noah Schiro
Julie Nethery Snyder
Mohammad Zarei

**Former employee: all work for this publication was done while at employer.*

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8596.iprd

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

**NIST Internal Report**
**NIST IR 8596 iprd**

# Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)

*NIST Community Profile*

Initial Preliminary Draft

Katerina Megas
Barbara Cuthill
*Applied Cybersecurity Division*
*Information Technology Laboratory*

Marissa Dotter
Michael Garris
Ishika Khemani
Bronwyn Patrick
Noah Schiro
Julie Nethery Snyder
Mohammad Zarei
*National Cybersecurity Center of Excellence*
*MITRE*

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8596.iprd

December 2025


U.S. Department of Commerce
*Howard Lutnick, Secretary*

National Institute of Standards and Technology
*Craig Burkhardt, Acting NIST Director and Acting Under Secretary of Commerce for Standards and Technology*

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

There may be references in this publication to other publications currently under development by NIST in accordance with its assigned statutory responsibilities. The information in this publication, including concepts and methodologies, may be used by federal agencies even before the completion of such companion publications. Thus, until each publication is completed, current requirements, guidelines, and procedures, where they exist, remain operative. For planning and transition purposes, federal agencies may wish to closely follow the development of these new publications by NIST.

Organizations are encouraged to review all draft publications during public comment periods and provide feedback to NIST. Many NIST cybersecurity publications, other than the ones noted above, are available at https://csrc.nist.gov/publications.

**NIST Technical Series Policies**
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

**How to Cite this NIST Technical Series Publication**
Megas K, Cuthill B, Snyder JN, Patrick B, Khemani I, Dotter M, Garris M, Zarei M, Schiro N (2025) Cybersecurity AI Profile: NIST Community Profile. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (IR) NIST IR 8596 iprd. https://doi.org/10.6028/NIST.8596.iprd

**Author ORCID iDs**
Katerina Megas: 0000-0002-2815-5448
Barbara Cuthill: 0000-0002-2588-6165
Julie Nethery Snyder: 0009-0004-6352-2831
Bronwyn Patrick: 0009-0001-7885-4773
Ishika Khemani: 0009-0002-0263-7374
Marissa Dotter: 0009-0003-8537-3091
Michael Garris: 0009-0008-0540-6719
Mohammad Zarei: 0009-0002-9104-3219
Noah Schiro: 0009-0001-7977-4089

**Public Comment Period**
December 16, 2025 – January 30, 2026

**Submit Comments**
cyberaiprofile@nist.gov

National Institute of Standards and Technology
Attn: Applied Security Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 2000) Gaithersburg, MD 20899-200

**Additional Information**
Additional information about this publication is available at https://csrc.nist.gov/pubs/ir/8596.iprd, including related content, potential updates, and document history.

**All comments are subject to release under the Freedom of Information Act (FOIA).**

## Abstract

The Cybersecurity Framework Profile for Artificial Intelligence (AI) Profile ("Cyber AI Profile" or "The Profile") will provide guidelines for managing cybersecurity risk related to AI systems as well as identifying opportunities for using AI to enhance cybersecurity capabilities. The Profile is organized using the NIST Cybersecurity Framework 2.0 outcomes (Functions, Categories, and Subcategories). This Preliminary Draft is shared along with a request for public comment to solicit feedback on the planned direction and content. Comments received will inform the initial public draft. More information about this project, including a roadmap, is available on the National Cybersecurity Center of Excellence (NCCoE) Cyber AI Profile project page.

## Keywords

artificial Intelligence; Community Profile; Cyber AI Profile; Cybersecurity Framework; risk management.

## Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof-of-concept implementations, and technical analyses to advance the development and productive use of information technology. ITL's responsibilities include the development of management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of other than national security-related information in federal information systems.

## Note to Reviewers

The purpose of this Preliminary Draft is to share insights regarding the direction of technical content in the Cyber AI Profile. NIST welcomes feedback and input on any aspect of this publication. Specifically, NIST seeks feedback on:

1. Document structure and topics:
    a) How do you envision using this publication? What changes would you like to see to increase/improve that use?
    b) How do you expect this publication to influence your future practices and processes?
    c) Are the proposed topics in this document sufficient to help your organization prioritize cybersecurity outcomes for AI?

2. Focus Area descriptions (Section 2.1):
    a) How well do the Focus Area descriptions reflect the scope and characteristics of AI usage? Are any characteristics missing, and if so, what are they and how should we describe them?
3. Profile content (Sections 2.3-2.8):
    a) When thinking about applying the Cyber AI Profile, how useful (or not) is it for all three Focus Areas to be shown alongside each other (as they are currently reflected)? What value might there be in providing Profile content for each Focus Area separately?
    b) What format(s) would be useful for providing the information in the Cyber AI Profile (e.g., a spreadsheet/workbook, the NIST Cybersecurity and Privacy Reference Tool (CPRT))?
    c) How well do the priorities and considerations discussed in Sections 2.3-2.8 relate to existing practices and standards leveraged by your organization? Are there significant gaps between current practices and those that are necessary to address unique characteristics of AI in each Focus Area that this publication should address? How should the AI-specific considerations inform the prioritization of each Subcategory?
    d) NIST published the Cybersecurity Framework (CSF) 2.0 Informative References and Implementation Examples to show potential ways to achieve the outcome in each Subcategory. This preliminary draft includes examples of Informative References for the Cyber AI Profile. Further literature review is in progress and NIST is seeking more input on Informative References to include. Which additional AI cybersecurity guidelines, standards, best practices, or mappings are you using that you recommend adding as Informative References for the Cyber AI Profile? For any Informative References you recommend, please share with us why you recommend them as well as how and why you would prioritize them for this document.
4. Glossary (Appendix B):
    a) NIST welcomes requests and suggestions for terms that should be added to this document's Glossary.

Commenters are encouraged to use the comment template provided on the NCCoE's Cyber AI Profile project page for responses to the question set and for specific comments on the text of the document. You may also submit your feedback and completed comment templates to the project team at cyberaiprofile@nist.gov. The deadline to submit comments is 11:59 p.m. Eastern Time on January 30, 2026.

All comments are subject to release under the Freedom of Information Act.

## Call for Patent Claims

This public review includes a call for information on essential patent claims (claims whose use would be required for compliance with the guidance or requirements in this Information Technology Laboratory (ITL) draft publication). Such guidance and/or requirements may be directly stated in this ITL Publication or by reference to another publication. This call also includes disclosure, where known, of the existence of pending U.S. or foreign patent applications relating to this ITL draft publication and of any relevant unexpired U.S. or foreign patents.

ITL may require from the patent holder, or a party authorized to make assurances on its behalf, in written or electronic form, either:

a) assurance in the form of a general disclaimer to the effect that such party does not hold and does not currently intend holding any essential patent claim(s); or

b) assurance that a license to such essential patent claim(s) will be made available to applicants desiring to utilize the license for the purpose of complying with the guidance or requirements in this ITL draft publication either:

   i. under reasonable terms and conditions that are demonstrably free of any unfair discrimination; or

   ii. without compensation and under reasonable terms and conditions that are demonstrably free of any unfair discrimination.

Such assurance shall indicate that the patent holder (or third party authorized to make assurances on its behalf) will include in any documents transferring ownership of patents subject to the assurance, provisions sufficient to ensure that the commitments in the assurance are binding on the transferee, and that the transferee will similarly include appropriate provisions in the event of future transfers with the goal of binding each successor-in-interest.

The assurance shall also indicate that it is intended to be binding on successors-in-interest regardless of whether such provisions are included in the relevant transfer documents.

Such statements should be addressed to: cyberaiprofile@nist.gov

## Table of Contents



## List of Tables

## List of Figures

## Acknowledgments

The authors extend their gratitude to all those who contributed insights that shaped discussions at the first Cyber AI Profile Workshop and the three Community of Interest (COI) Working Sessions. Thank you to the many stakeholders across sectors, and the participants, speakers, and moderators of the workshop and working sessions, as well as our colleagues at NIST and The MITRE Corporation who offered the expertise necessary to conduct these events and review this draft. The thoughtful feedback shared by stakeholders on the initial concept paper and through our on-going engagement was instrumental in shaping this Preliminary Draft of the Cyber AI Profile.

## Executive Summary

Recent advancements in Artificial Intelligence (AI) technology introduce new cybersecurity opportunities and challenges for organizations. Long active in both cybersecurity and AI, NIST has worked closely with stakeholders in conducting research and advancing standards and technology. NIST is utilizing its cybersecurity and AI resources to assist organizations in addressing these new AI opportunities and challenges.

Working with the cybersecurity and AI communities, NIST is creating the Cybersecurity Framework Profile for Artificial Intelligence Profile ("Cyber AI Profile" or "The Profile") to create a consistent and common approach for discussing how AI advances inform organizations' cybersecurity risk management strategies, as well as to provide a resource for organizations to strategically adopt AI while addressing and prioritizing emerging cybersecurity risks stemming from advancements in AI. The Cyber AI Profile is not meant to replace existing frameworks, but provides prioritized cybersecurity guidelines for organizations securing AI, using AI to enhance cybersecurity defenses, or defending against adversarial uses of AI. The Cyber AI Profile will use the CSF 2.0 Functions, Categories, and Subcategories to provide informed, structured, technology-neutral recommendations for cybersecurity and AI professionals.

The Cyber AI Profile will address the following three Focus Areas:

- **Securing AI System Components (Secure):** Focuses on identifying cybersecurity challenges when integrating AI into organizational ecosystems and infrastructure.
- **Conducting AI-Enabled Cyber Defense (Defend):** Focuses on identifying opportunities to use AI to enhance cybersecurity processes and activities, and understanding challenges when leveraging AI to support defensive operations.
- **Thwarting AI-Enabled Cyber Attacks (Thwart):** Focuses on building resilience to protect against new AI-enabled threat vectors.

This document is a Preliminary Draft of the Cyber AI Profile. The Preliminary Draft is intended to convey current thinking regarding the direction of the technical content, identify additional Informative References that the document should include, and to seek feedback from the public to inform the Initial Public Draft.

## 1. Introduction

The National Institute of Standards and Technology (NIST) Cybersecurity Framework (CSF) 2.0 provides voluntary guidance to help organizations manage, reduce, and communicate cybersecurity risks. The Framework was created through collaboration with a wide range of stakeholders and encourages a prioritized, risk-based approach to addressing cybersecurity issues. The NIST Artificial Intelligence Risk Management Framework (AI RMF) broadly addresses the responsible use of Artificial Intelligence (AI) systems and points to the NIST CSF and NIST Risk Management Framework (RMF) as some of the available tools for managing any associated cybersecurity risks.

Although the Frameworks present a variety of mitigations, Community Profiles apply the CSF to prioritize cybersecurity outcomes in a specific context. The Cybersecurity Framework Profile for Artificial Intelligence (AI) Profile ("Cyber AI Profile" or "The Profile") serves this purpose for AI. NIST is also in the process of developing a series of NIST Special Publication (SP) 800-53 Control Overlays for Securing AI Systems (COSAiS) as implementation-focused guidelines to help organizations customize and prioritize the most critical controls to consider when using AI systems.

To develop this Cyber AI Profile, the NIST engaged stakeholders from government, academia, and industry to better understand the current state of the cybersecurity challenges facing the AI and cybersecurity communities. NIST first published the Cybersecurity and AI Workshop Concept Paper in February 2025 to gather stakeholder perspectives on a CSF Community Profile based on the intersection of cybersecurity and AI. These perspectives were discussed in a public workshop held by NIST in April 2025. In April 2025, NIST conducted a public workshop focused on the benefits and challenges associated with leveraging AI in organizational and cybersecurity operations. This workshop also focused on how organizations can leverage existing NIST Frameworks, such as the CSF, RMF, AI RMF, and the Privacy Framework to help address the identified challenges. Participants also explored the scope and practical applications of the Cyber AI Profile. NIST published reflections from the April 2025 workshop to document the concepts identified in the workshop, including the benefits, concerns, and challenges associated with adopting AI, the current state of relevant cybersecurity risk management practices [1], gaps in securely implementing and leveraging AI for organizational operations, as well as thwarting AI-enabled cybersecurity attacks.

As a follow-on effort, in August and September of 2025, NIST conducted three public working sessions designed specifically to identify cybersecurity priorities needed to develop the Cyber AI Profile. Each working session explored one of the three Focus Areas, with the intention of learning more from the AI and cybersecurity communities about what support cybersecurity programs need as they manage the impacts of AI to their organization. Materials from these working sessions can be found on each event page under the Post Event Materials section (Conducting AI-enabled Cyber Defense, Securing AI System Components, and Thwarting AI-enabled Cyber Attacks).

As a Preliminary Draft, this Profile is still in development.

## 1.1. Purpose

In discussions with members of the AI and cybersecurity communities, it was strongly suggested that there would be value in developing guidance based on the CSF to address the cybersecurity challenges related to AI adoption and use. Organizations may vary on whether and how they are using AI within their operations—some may not yet be using AI, and some may only be using cybersecurity solutions enabled with basic machine learning (ML) technology for pattern detection and planning, and some may be exploring newer and transformational AI capabilities, such as Generative AI (GenAI).

However, regardless of where organizations are on their AI journey, their cybersecurity programs need risk management approaches that support and integrate the realities of advancements in AI. The Cyber AI Profile is intended to support cybersecurity programs as they manage the impacts of advancements in AI to their organization by helping organizations secure AI system components, such as models, agents, algorithms, prompts, and data; take advantage of the new opportunities AI offers to improve cybersecurity defenses; and prepare for changes to the threat landscape based on adversarial use of AI.

In doing so, the Profile seeks to help organizations integrate AI-specific considerations into existing cybersecurity programs by:

- Establishing a shared understanding of AI-related cybersecurity priorities and considerations for any organization;
- Fostering collaboration and communication across the AI and cybersecurity communities; and
- Providing common AI cybersecurity target outcomes that organizations can use to support strategic planning efforts and cybersecurity assessments, based on the CSF Core.

The Cyber AI Profile provides information that organizations can integrate into existing cybersecurity programs to manage AI opportunities and risk alongside other systems, data and technologies. Like any new technology, implementing AI can introduce new risks. AI can make mistakes, offer confident yet incorrect answers, leak sensitive data, and be manipulated by attackers. And, just like other types of systems, organizations strive for AI systems that are trustworthy and support organizational objectives for demonstrating responsibility and accountability. In addition to the feedback on the Concept Paper, the public workshop, and the three COI working sessions, an on-going and in-depth literature review informed the creation of this Profile. This Profile builds on and highlights existing work in the cybersecurity and AI fields useful to organizations as they endeavor to update their cybersecurity risk management approaches. The Cyber AI Profile also aims to facilitate collaboration and mutual understanding between the cybersecurity and AI communities to address shared risks and opportunities.

## 1.2. Scope

The term Artificial Intelligence has a long history with broad connotations and many different meanings depending on when the term is being used and other contextual factors. Due to the

rapidly evolving nature of the field, and an effort to create a publication which is useful and flexible long into the future, this Profile does not assert a definition of the term "AI" and leaves application of this Profile open to the broadest sense of the term[1].

The Profile does, however, offer examples of AI throughout to facilitate understanding and use of this document and uses the term "AI systems" throughout the Profile to refer to any systems that are using AI capabilities, whether they are stand-alone AI systems or applications, infrastructure, and organizations that incorporate AI. Examples of types of AI systems in common use include, but are not limited to:

- Large Language Models (LLMs) for understanding, interacting with, generating, and summarizing language-based content.
- Generative AI systems for content such as text, code, images, video, and audio.
- Domain-specific optimization systems such as for scheduling or load balancing.
- Prediction and anomaly detection systems.
- Expert systems.
- Data mining, "big data," and recommendation systems.
- Search engines.
- Automated and agentic systems.

AI systems may be based on algorithms and techniques that include, but are not limited to:

- Statistical machine learning (such as regression, clustering, genetic algorithms, decision trees, and deep learning neural networks).
- Logic and symbolic reasoning (such as inductive logic programming, human-created heuristic rules, and fuzzy logic).
- Heuristic search and planning (such as A* search).
- Reinforcement and apprentice learning (such as reinforcement learning with human feedback (RLHF)).
- Ensemble, hybrid, and neuro-symbolic techniques.

There are many cybersecurity considerations which impact all these systems, algorithms, and techniques. In many cases, there may be specific types of AI which have unique considerations. These will be highlighted throughout the Cyber AI Profile.

The discussions in this Cyber AI Profile are oriented around three Focus Areas:

- Securing AI System Components (Secure)

[1] NIST's AI RMF "refers to an AI system as an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022)" The specific types of AI systems are intentionally left broad. There are many types of AI, each with their own set of considerations, mitigations, and response strategies. This document covers broad considerations that encompass most types of AI systems to maintain relevance in the near-term evolution of the field.

- Conducting AI-Enabled Cyber Defense (Defend)
- Thwarting AI-Enabled Cyber Attacks (Thwart)

This Profile organizes discussions about the cybersecurity implications for these three Focus Areas around the Functions, Categories, and Subcategories of the CSF Core. Considerations beyond cybersecurity are outside the scope of the Cyber AI Profile.

### 1.3. Audience

The Cyber AI Profile is intended for any organization that:

- Is using AI technologies, whether they are stand-alone AI systems (e.g., chatbots and document summarizers) or AI-enabled capabilities that are integrated into other systems (e.g., AI for analyzing and balancing demand across a power grid), regardless of whether the organization builds or purchases AI technologies;
- Would like to understand and capitalize on the cybersecurity capabilities AI can provide (e.g., AI analytics in cybersecurity tools);
- Would like to better understand and defend against AI-enabled cyber-attacks (e.g., against adversaries using AI to generate convincing malicious emails); and
- Is developing AI systems.

This Cyber AI Profile can be used by organizations to identify and communicate cybersecurity expectations regarding AI with internal and external stakeholders. The Profile can also be used by organizational leadership to generate priorities tailored to the operational aspects of the organization.

To complement this Cyber AI Profile and support the adoption of the AI Risk Management Framework, NIST is developing a series of *Control Overlays for Securing AI Systems (COSAiS)* using the NIST Special Publication (SP) 800-53 controls. Control overlays enable organizations and communities of interest to tailor the controls (or control baselines) for their specific context and needs. COSAiS will provide additional implementation-focused guidelines and assist users and developers of AI systems manage unique risks in different use cases:

- Adapting & Using Generative AI
- Using & Fine-Tuning Predictive AI
- Using Agentic AI: Single Agent & Multi-Agent

Learn more about the Control Overlays for AI Project, Slack space, and how to join the Slack channel at: https://csrc.nist.gov/projects/cosais
Questions and comments can be directed to overlays-securing-ai@list.nist.gov.

### 1.4. Document Structure

The remainder of this document is organized into the following sections:

- Section 2 discusses Focus Areas and provides the proposed prioritized Subcategories and considerations for each Focus Area, which comprises the Cyber AI Profile.
- Section 3 contains cited resources.

- Appendix A contains a list of symbols, abbreviations, and acronyms used in the Profile.
- Appendix B will contain a glossary in future drafts.
- Appendix C contains an overview of the NIST CSF.
- Appendix D provides ideas for using the Cyber AI Profile

## 2. The Cyber AI Profile

**NOTE:** This section may expand in the IPD if NIST receives feedback that more information is needed to explain the Focus Areas.

The Cyber AI Profile is organized by the elements of the CSF (Functions, Categories, and Subcategories) and Focus Areas. The Profile assumes organizations already have a cybersecurity program in place and provides additional considerations based on an organization's deployment and use of AI systems. Content in the Cyber AI Profile was developed based on literature review, inputs from subject matter experts, and working sessions held with stakeholders and the public.

### 2.1. Focus Areas

The intersection of cybersecurity and AI has multiple facets. To provide an organizing construct the Cyber AI Profile offers three Focus Areas that are based on inputs from stakeholders, each with a single keyword for ease of reference in the detailed Cyber AI Profile discussions. The Focus Areas are:

- **Securing AI System Components (Secure)**: Focuses on managing cybersecurity challenges when integrating AI into organizational ecosystems and infrastructure.
- **Conducting AI-Enabled Cyber Defense (Defend):** Focuses on identifying opportunities to use AI to enhance cybersecurity processes and activities, and understanding challenges when leveraging AI to support defensive operations.
- **Thwarting AI-Enabled Cyber Attacks (Thwart):** Focuses on building resilience to protect against new AI-enabled threat vectors.

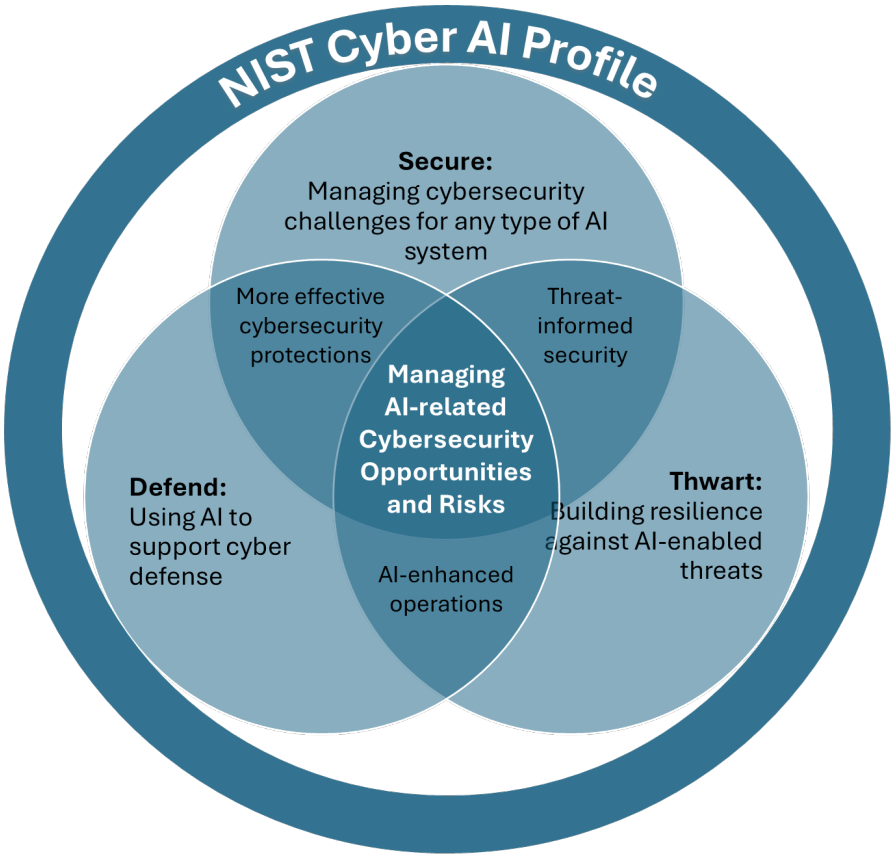
While the three Focus Areas each address AI-related cybersecurity risk from a different angle, they share some commonality, and each Focus Area enables the other two. The paragraphs that follow provide an overview of these distinctions and relationships, which are summarized in Fig. 1 below.

***Secure*** focuses on managing cybersecurity challenges when integrating any type of AI system into an organization's environment, while considering how an organization's cybersecurity program may need to adapt to accommodate the unique needs of AI systems. ***Defend*** focuses on how organizations can capitalize on the opportunities AI offers to improve their cybersecurity defensive capabilities, such as taking advantage of new efficiencies. While ***Secure*** considers the cybersecurity needs for integrating and managing a broad spectrum of AI systems, ***Defend*** considers how AI helps organizations better and more proactively conduct cybersecurity activities. Both Focus Areas often work hand in hand—when AI systems are well secured, AI-enabled defenses are better able to find and react to threats targeting those systems.

***Secure*** enables an organization's foundational cybersecurity practices, which integrate with resilience measures to prevent AI-enabled attack addressed by ***Thwart***. ***Thwart*** acknowledges that AI can play a significant role in helping adversaries become more sophisticated in their

attack methods and carry out attacks more efficiently. The knowledge regarding adversarial use of AI from ***Thwart*** in turn provides insights that help an organization build resiliency into its overall cybersecurity program under ***Secure***.

Together, the cybersecurity considerations for these three Focus Areas Help organizational cybersecurity programs ***Secure*** their AI systems, take advantage of new AI capabilities to ***Defend*** their organizations, and proactively ***Thwart*** AI-enabled attacks, and strengthen cybersecurity defenses based on known threats.




405 **Fig. 11: Relationship Between Cyber AI Profile Focus Areas**

406 The scope and characteristics of each Focus Area are described in the subsections that follow.
407 The Focus Area descriptions informed the decisions made when developing the Cyber AI Profile
408 content in Sections 2.3-2.8.

409 ### 2.1.1. Securing AI System Components (Secure)

410 The Secure Focus Area supplements existing cybersecurity and risk management best practices
411 to address novel, expanded, and altered attack surfaces associated with the integration of AI
412 systems into an organization, its ecosystems, and its infrastructure. This covers the AI systems
413 themselves, their supply chains, including data and machine learning infrastructure, and the
414 other systems and data that the AI system relies on.

415 AI systems are becoming increasingly important to day-to-day business operations in many
416 sectors. Examples of uses of AI that fall within the scope of Secure include, but are not limited
417 to:

418 • Restaurants to forecast demand and take/make orders.
419 • Insurance companies to determine risk and premiums.
420 • Power grids to balance loads.

- Companies and individuals to filter, prioritize, summarize, generate, and proofread emails, reports, and other documents.
- Customer service organizations to perform initial interactions with customers.

Each of these sectors and users may have methods for handling cybersecurity challenges but are still seeking to understand how AI-augmented systems change these cybersecurity risks. Compared to other types of computer systems, AI behavior and vulnerabilities tend to be more contextual, dynamic, opaque, and harder to predict, as well as more difficult to identify, verify, diagnose, and document, when they appear. Once issues are identified, they can also be difficult to mitigate as some vulnerabilities are inherent to the AI model or the underlying training data and machine-learning infrastructure.

AI presents organizations with opportunities to improve cybersecurity processes while also introducing new considerations. Each organization will need to determine how to appropriately assess and manage the impacts of adopting and using AI. Cybersecurity practitioners need to implement a range of measures related to AI systems, including addressing AI considerations in employee training, securing the AI systems from adversarial inputs, and maintaining data quality and integrity.

### 2.1.2. Conducting AI-Enabled Cyber Defense (Defend)

The scope of the Defend Focus Area is identifying opportunities for the use of AI in supporting cybersecurity processes and activities and understanding the challenges that come with leveraging AI to assist in defensive operations. AI is increasingly influential and important to organizational cybersecurity defense— for example, it can support cybersecurity teams by sifting through the significant volume of alerts, distinguishing actual threats from noise, distinguishing the most serious threats, and even prioritizing and providing suggestions about what actions to take during an active attack. Some teams are also experimenting with agentic AI, where multiple AI agents coordinate on the identification of attacks and take actions to defend against them while running checks on each other to improve defenses. Examples of opportunities to use AI to enhance cyber defense capabilities include:

- Mission assurance:
    - Security governance and policy (e.g., ensuring AI outputs follow organizational policies)
    - Configuration management (e.g., managing configuration drift)
    - Supporting automated policy enforcement and compliance assessments to cybersecurity regulations and standards, including streamlining compliance tasks[2]

[2] Examples include: NIST 800-171, ISO 31000, Defense Federal Acquisition Regulation Supplement (DFARS)/Cybersecurity Maturity Model Certification (CMMC), capability maturity model integration (CMMI)/Federal Risk and Authorization Management Program (FedRAMP).

- Enhance and accelerate information sharing (e.g., use of Structured Threat Information eXpression (STIX)/Open Cyber Threat Intelligence (OpenCTI) formats for interoperability and cross-organization sharing)
- Risk communication and executive decision support (e.g., helping CISOs translate risk to business language)
- Recovery planning (e.g., prioritize which systems to restore and generate communications during recovery)

- Predictive and proactive:
  - Proactive risk management (e.g., predictive analysis of cyber attacks)
  - Identify potential threat actors and tactics that may exploit vulnerabilities (e.g., via open-source intelligence (OSINT) datasets for advanced persistent threat (APT) profiling and integration with ATT&CK®/cyber threat intelligence (CTI) tools)
  - Predictive maintenance and asset risk forecasting (e.g., tracking asset lifecycle, predicting end-of-life (EOL) risk, and plan maintenance)
  - Agent-based AI defense (e.g., “swarm” agents, agent sanity check, agentic AI security reference architectures, continuous validation and authentication with digital credentials)
- Investigation and analysis:
  - Advanced threat detection and analysis (e.g., anomaly detection, outlier detection, and user and entity behavior analytics (UEBA)) to identify and reduce false positive (FP) and false negative (FN) rates, identify insider threats, and prevent fraud)
  - Adapting zero trust modeling to new threat patterns by detecting model drift. (e.g., source code vulnerabilities detection using AI, using LLM to interpret the semantic and logic of source code to identify malicious behavior, continuously monitoring and verifying user behavior and system activity.)
- Response and remediation:
  - Adversarial training and simulation (e.g., provide realistic scenarios to train cybersecurity personnel, real time coaching to improve incident response in real time, adaptive learning to adjust scenarios)
  - Automated incident response (e.g., AI can help with automated lock down, incident playbook, and consistency in incident responses)
  - Creating efficiencies for cybersecurity help desks (e.g., handling first-tier requests, prioritizing reports)
  - Conduct reporting (e.g., writing incident reports and client monitoring outputs)

AI can improve defensive processes by augmenting human analysts, enhancing detection and response time, and supporting recovery. Using AI for cybersecurity defense is a dynamic area and organizations will need to continuously evaluate whether capabilities are sufficiently mature for their needs. These opportunities for using AI to bolster cybersecurity defenses are contemplated in the Profile, as well as new risks that may arise from using AI in these ways.

### 2.1.3. Thwarting AI-Enabled Cyber Attacks (Thwart)

While AI-enabled attacks are increasingly being reported, many will likely go undetected until adversary use of AI is better understood, and effective measures are integrated into cyber defenses. The Thwart Focus Area addresses how AI can enhance adversary capabilities, how these attacks could impact the entire cybersecurity landscape, and what organizations can do to bolster their systems against these emerging threats.

As cyber attacks evolve and AI becomes increasingly more capable in offensive cyber-related operations [11, 12, 13, 14], it is necessary to build resilience and robustness into systems to scale with the growing threat. As with other types of cybersecurity attacks, AI-enabled cyber attacks could support adversaries across all aspects of the cybersecurity landscape, including discovering exploitable vulnerabilities or weaknesses, advancing attack paths in shorter timelines, exfiltrating and tampering with data, and uplifting previously unsophisticated bad actors [11, 12, 13]. What differentiates these AI-enabled attacks other types of attacks is the:

- speed and scale with which the attacks occur, making countermeasures harder to implement within typical timelines;
- ease with which adversaries could deploy AI-enabled attacks; and
- dynamic and optimized nature of AI.

These factors will have an impact on the ability to properly and efficiently identify, mitigate, and establish defenses for this evolving threat landscape [14].

Attacks targeting personnel are one of the leading ways AI-enabled cyber-attacks are gaining a foothold into systems. AI-enabled spear-phishing attacks exploit users through more realistic communications including audio and video manipulation, such as with DeepFake attacks. Additionally, new Generative AI (GenAI) techniques can create hyper-realistic malicious websites and links that can be spread through email phishing. These emerging AI-enabled cyber attacks take little effort on the adversary to generate, and with the proliferation of personal data distributed on social media and online, they can create personalized profiles of their intended targets that follow trust-building narratives. These attacks showcase the need for updated training for personnel and integrated, automated defenses to bolster current email and authentication security measures.

Another example of AI-enabled attacks comes from adversaries leveraging AI to generate new forms of malware to obfuscate their intentions or make current signature-based detections and

anti-virus systems fail, such as malware that is executed from computer memory rather than hardware and unlikely or seemingly benign file types to hide planted malware instructions.

A third example of AI-enabled attacks involves the use of AI agents to autonomously orchestrate different phases of a cyber attack, from reconnaissance and attack surface mapping to vulnerability exploitation to credential harvesting, lateral movement, and data collection. AI agents are increasingly capable of autonomously operating common cybersecurity tools and utilities, such as network scanners, password crackers, exploitation frameworks, and binary analysis suites [14].

The Thwart Focus Area builds on capabilities under Secure and Defend to draw out additional proactive measures to get out in front of adversarial uses of AI.

## 2.2. How to Read the Cyber AI Profile

Tables 1-6 use the CSF Core to summarize AI-related considerations, where relevant, for each of the 106 CSF Subcategories3. Each table addresses a single CSF Function. Each table contains the following columns:

- **CSF Core**: provides the CSF Core elements (Functions, Categories, and Subcategories) and their descriptions.
- **General Considerations:** contain non-exhaustive AI-related considerations for achieving the outcomes in each Subcategory (where relevant) that apply across more than one Focus Area and that can inform an organization's prioritization of a Subcategory.
- **Focus Area Proposed Priorities & Considerations:** includes a sub-column for each of the three Focus Areas, which includes:
    - A **Proposed Priority** (described below) for indicating how important it may be for organization to achieve the Subcategory's outcome(s) to meet the objective of each Focus Area.
    - **Sample Opportunities** are included only for the Defend Focus Area to present opportunities to leverage AI to help achieve the outcomes in the Subcategory. Sample Opportunities are not provided for every Subcategory since not every Subcategory will have an AI opportunity.
    - Non-exhaustive AI-related **Sample Focus Area Considerations** to help organizations effectively achieve the outcomes in the Subcategory (where relevant to the objective of the Focus Area). The considerations supplement existing cybersecurity practices with the intent to integrate AI considerations into existing programs. Considerations are based on observations in the field and/or subject matter expertise.
      The phrase "standard cybersecurity practices apply" is used to indicate that there are no unique considerations identified for the Focus Area and the

[3] Considerations only appear where there are unique AI-related concepts for implementing the Subcategory.

activities of the cybersecurity program are sufficient for AI systems. These considerations also include the rationale for Subcategories that are designated as a High (1) or Moderate (2) proposed priority. For Defend, these considerations are only related to the opportunities listed.

- **Example Informative References** includes a broad range of widely available and broadly applicable resources (laws, standards, guidelines and other research publications). These supplementary materials informed the Subcategory considerations and proposed priorities. They may also be useful for organizations as they consider how to achieve the outcomes in the Subcategory. Organizations can also use the Example Informative References to understand where they may already be achieving those outcomes. (**NOTE:** The NIST SP 800-53 controls in the General Considerations column are listed exactly as they appear in the crosswalk available through the NIST National Online Informative Reference (OLIR) Catalog[4]).[5]

Determining the considerations and priority of each Subcategory is a subjective exercise that is based on observations in the field and/or subject matter expertise. The Cyber AI Profile offers these proposed priorities to help organizations determine which Subcategories they may wish to address sooner. The priorities are not intended to reflect the degree of difficulty in achieving the Subcategory. The priority level of Subcategories may be higher or lower for individual organizations based on characteristics of the environment, needs, risk tolerance, or other factors. The proposed level of Subcategory priorities is indicated in each Table by number:

- **"1" for High Priority:** These Subcategories are considered the most critical to address the challenges for a Focus Area. High priority Subcategories should typically be addressed most immediately given available resources.
- **"2" for Moderate Priority:** These Subcategories should be the next priority after implementing High Priority Subcategories.
- **"3" for Foundational Priority:** These Subcategories are generally important to the Focus Area but may not require the same level of urgency as higher priorities. Note that "Foundational" does not equate to low priority. All Subcategories should receive consideration.

Although organizations should develop strategies that address all Subcategories as part of a cybersecurity program, the prioritizations offered in the Cyber AI Profile provide adaptable guidance that suggests cybersecurity capabilities that will provide the greatest impact toward meeting the AI-related challenges associated with each Focus Area. As in any risk management discussion the decision to deploy AI-related cybersecurity mitigations should be evaluated in

[4] The mapping between CSF 2.0 and NIST SP 800-53, Rev 5, as available at: https://csrc.nist.gov/projects/olir/informative-reference-catalog/details?referenceId=131#/ (last accessed 12/5/2025).
[5] Per the AI Action Plan, the AI RMF is currently in revision and will be included in a future version.

598 the context of the organization's needs and risk tolerances. Trade-offs may vary for a given
599 mitigation in different environments.

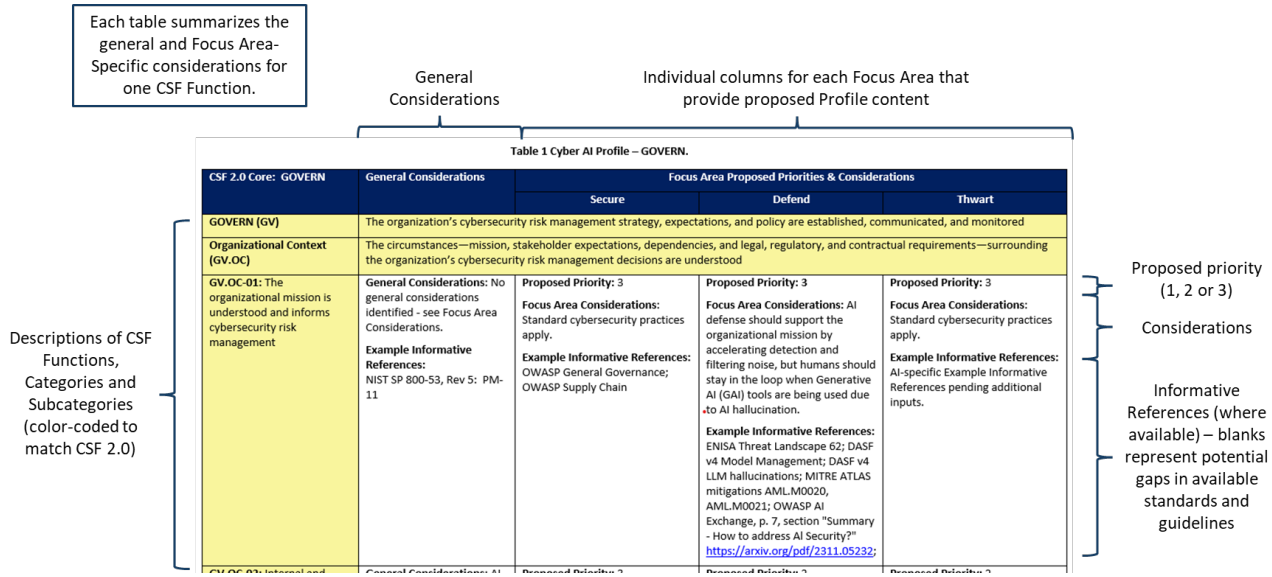



**Table 1 Cyber AI Profile – GOVERN.**

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **GOVERN (GV)** | The organization's cybersecurity risk management strategy, expectations, and policy are established, communicated, and monitored | | | |
| **Organizational Context (GV.OC)** | The circumstances—mission, stakeholder expectations, dependencies, and legal, regulatory, and contractual requirements—surrounding the organization's cybersecurity risk management decisions are understood | | | |
| **GV.OC-01:** The organizational mission is understood and informs cybersecurity risk management | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-11 | **Proposed Priority:** 3<br>**Focus Area Considerations:** Standard cybersecurity practices apply.<br>**Example Informative References:** OWASP General Governance; OWASP Supply Chain | **Proposed Priority: 3**<br>**Focus Area Considerations:** AI defense should support the organizational mission by accelerating detection and filtering noise, but humans should stay in the loop when Generative AI (GAI) tools are being used due to AI hallucination.<br>**Example Informative References:** ENISA Threat Landscape 62; DASF v4 Model Management; DASF v4 LLM hallucinations; MITRE ATLAS mitigations AML.M0020, AML.M0021; OWASP AI Exchange, p. 7, section "Summary - How to address AI Security?" https://arxiv.org/pdf/2311.05232; | **Proposed Priority:** 3<br>**Focus Area Considerations:** Standard cybersecurity practices apply.<br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

600 **Fig. 2: How to Read the Cyber AI Profile Preliminary Draft Tables**

601 **NOTE:** Work remains ongoing to gather input and feedback, and significant changes are
602 possible before finalizing the Cyber AI Profile. The Cyber AI Profile Preliminary Draft and the
603 tables that follow are an early representation of work in progress and we look forward to your
604 input regarding additional information to include.

605 This preliminary draft reflects examples of Informative References. Further literature review is
606 in progress and NIST is seeking more input on Informative References that should be included
607 in the Cyber AI Profiles to help us strengthen the considerations.

608 For ease of providing all the detailed Cyber AI Profile content in one place, the Preliminary Draft
609 appears in static tables that can fit on a standard size printed page. NIST is interested in
610 feedback regarding more usable formats for sharing the contents of the Cyber AI Profile to
611 enhance its usability.

## 2.3. Cyber AI Profile: GOVERN

**Table 1 Cyber AI Profile – GOVERN.**

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **GOVERN (GV)** | The organization's cybersecurity risk management strategy, expectations, and policy are established, communicated, and monitored | | | |
| **Organizational Context (GV.OC)** | The circumstances—mission, stakeholder expectations, dependencies, and legal, regulatory, and contractual requirements—surrounding the organization's cybersecurity risk management decisions are understood | | | |
| **GV.OC-01:** The organizational mission is understood and informs cybersecurity risk management | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-11 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain | **Proposed Priority:** 3<br><br>**Sample Opportunities:** Standard cybersecurity practices apply.<br><br>**Sample Focus Area Considerations:**<br><br>Standard cybersecurity practices apply.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 50; ATLAS AML.M0020; OWASP AI Exchange: AI Security Overview https://arxiv.org/pdf/2311.05232; NIST AI 100-2e2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **GV.OC-02:** Internal and external stakeholders are understood, and their needs and expectations regarding cybersecurity risk management are understood and considered | **General Considerations**: AI use introduces considerations from multiple aspects of organizational operations, including legal, technical, procurement/acquisitions, and governance teams. Collaboration across these areas is essential for addressing AI-related | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 13, 40, 51, 53; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Understanding the strengths and limitations of AI capabilities for cyber defense is important for meeting stakeholder expectations and to ensure the balance between the | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | cybersecurity risks. Multidisciplinary approaches facilitate a comprehensive enterprise view.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-18; PM-30; SR-03; SR-05; SR-06; SR-08; Reflections from the First Cyber AI Profile Workshop | LLM03 Supply Chain; ENISA Threat Landscape 2025 | required human oversight and automation.<br><br>**Example Informative References:** DASF 38,50;OWASP AI Exchange: AI Transparency; ENISA Threat Landscape 2025; ATLAS AML.M0003 | |
| **GV.OC-03:** Legal, regulatory, and contractual requirements regarding cybersecurity—including privacy and civil liberties obligations—are understood and managed | **General Considerations**: The AI legal, regulatory, and standards landscape is rapidly evolving and will impact the decisions organizations make regarding whether, when, and how to use AI. Organizations need measures in place maintain awareness of their responsibilities. Human oversight will be required to maintain regulatory and legal compliance.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; PM-28; PT | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** The legal framework regarding AI usage is evolving, particularly in areas like cybersecurity, privacy, fair use of copyrighted material, and AI training.<br><br>**Example Informative References:** DASF 32, 40; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional runtime controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority: 1**<br><br>**Sample Opportunities:**<br><br>AI capabilities can support compliance with legal, regulatory, and contractual requirements by analyzing and summarizing requirements, accelerating policy development, speeding up review processes, identifying and mapping similar concepts between documents, and even simplifying audit processes by using automation to monitor, identify, and correct noncompliance issues in real time.<br><br>**Sample Focus Area Considerations:** Defensive AI tools handle logs and sensitive data in line with privacy obligations including consent, | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | usage and aggregation controls and new AI specific laws.<br><br>AI audits are designed to demonstrate compliance with legal, regulatory, and contractual requirements, while addressing AI specific needs like explainability.<br><br>**Example Informative References:** DASF 44,50; ENISA Threat Landscape 2025 ; ATLAS AML.M0005 | |
| **GV.OC-04:** Critical objectives, capabilities, and services that external stakeholders depend on or expect from the organization are understood and communicated | **General Considerations**: AI impacts on cybersecurity are managed. The cybersecurity capabilities and limitations of AI are understood and communicated to the users. Guardrails and backup plans are established and implemented.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-08; PM-11; CP-02(08); PM-30(01); RA-09 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Communicating the intended use and known limitations of AI is critical to effective usage. Understanding how it makes decisions, when it makes mistakes, and what to do when it makes mistakes are communicated to users.<br><br>**Example Informative References:** DASF 11, 19, 40-42, 51, 53; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029 ; | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Organizations to understand and communicate, how AI-enabled cybersecurity defense support services that stakeholders depend on including defensive decision support and AI detection.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 50; OWASP AI Exchange: Oversight; ATLAS AML.M0008; https://arxiv.org/pdf/2309.01029 ; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/pdf/2309.01029 ; https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | https://arxiv.org/html/2503.11917v3 | | |
| **GV.OC-05:** Outcomes, capabilities, and services that the organization depends on are understood and communicated | **General Considerations**: AI output is unpredictable. The capabilities and limitations of a system are understood and communicated to the users. Guardrails and backup plans are established and implemented.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-11; PM-30; RA-07; SA-09; SR-05 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Organizations identify business outcomes that rely on AI systems and clearly communicate these dependencies to relevant teams. Additionally, understanding the intended use and limitations of an AI model is critical to effective usage and cybersecurity measures. Users need to understand how AI makes decisions, how to identify when it makes mistakes, and what to do when it makes mistakes.<br><br>**Example Informative References:** DASF 23, 45, 51, 53; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Organizations should identify which defense capabilities rely on AI systems and clearly communicate these uses and dependencies to relevant teams including. Safeguarding AI defense actions against adversarial manipulation, model drift, and hallucination with HITL checks and confidence thresholds that indicate whether an AI output is reliable enough to act upon.<br><br>**Example Informative References:** OWASP AI Exchange: Supply Chain Management; DASF 39,50; ATLAS AML.M0023; https://arxiv.org/pdf/2309.01029 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/pdf/2309.01029 |
| **Risk Management Strategy (GV.RM)** | The organization's priorities, constraints, risk tolerance and appetite statements, and assumptions are established, communicated, and used to support operational risk decisions | | | |
| **GV.RM-01:** Risk management objectives are established and agreed to by organizational stakeholders | **General Considerations:** Evaluate how AI use aligns with cybersecurity risk management objectives. Include AI specific goals when establishing cybersecurity risk management | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI analyzes vulnerabilities to issue early warnings and helps prioritize | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | objectives (e.g., FP reduction, improving triage speed).<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; RA-07; SR-02 | **Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | defensive actions before incidents occur.<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 50; ATLAS AML.M0000 | **Example Informative References:** https://arxiv.org/html/2503.11917v3 |
| **GV.RM-02:** Risk appetite and risk tolerance statements are established, communicated, and maintained | **General Considerations**: The risk tolerance for AI needs to be reevaluated on a frequent basis, due to the evolving nature of threat and defense capabilities.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Integrate AI-specific risks into the organization's formal risk appetite and tolerance statements. Update risk appetite and tolerance on a regular basis as the nature of AI systems evolve.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Integrate AI-specific risks into the organization's formal risk appetite and tolerance statements. Update risk appetite and tolerance on a regular basis as the nature of AI systems evolve.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025; ATLAS AML.M0019 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** New risk tolerance recommendations may be needed with emerging AI-enabled threats and attacks.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |
| **GV.RM-03:** Cybersecurity risk management activities and outcomes are included in enterprise risk management processes | **General Considerations:** Integrate AI risk management into existing enterprise risk management practices and governance structures.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-03; PM-09; PM-30; RA-07; SA-24; SR- | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 5, 13; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Ensure that AI security threats and the effectiveness of AI-specific security controls (e.g. data provenance tracking) are formally reported and factored into the | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | 02; Reflections from the First Cyber AI Profile Workshop | Landscape 2025; https://arxiv.org/html/2503.11917v3 | organization's enterprise risk management.<br><br>**Example Informative References:** DASF 38,50; ENISA Threat Landscape 2025; ATLAS AML.M0023; https://arxiv.org/html/2503.11917v3 | |
| **GV.RM-04:** Strategic direction that describes appropriate risk response options is established and communicated | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-28; PM-30; SR-02 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI analyzes vulnerabilities to issue early warnings and helps prioritize defensive actions before incidents occur**.**<br><br>AI translates complex technical data into clear business language by summarizing technical risks into concise insights and helping executives make informed risk decisions.<br><br>**Sample Focus Area Considerations:** Establish and communicate specific strategic risk response options (e.g. mitigations, avoidance) tailored for unique AI-related risks, such as prompt injection, model inversion, and data poisoning.<br><br>**Example Informative References:** OWASP AI Exchange: Risk Treatment; DASF 39; ENISA Threat | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | Landscape 2025; ATLAS AML.M0011; https://arxiv.org/html/2503.11917v3 | |
| **GV.RM-05:** Lines of communication across the organization are established for cybersecurity risks, including risks from suppliers and other third parties | **General Considerations**: Create AI-specific communication channels for sharing information about threat vectors and/or defense capabilities.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-30 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI is a rapidly evolving space that is likely to require frequent updates regarding cybersecurity risks.<br><br>**Example Informative References:** DASF 32, 51, 53; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI formats and shares threat intelligence using standard protocols (such as STIX and OpenCTI) to ensure teams and partners stay aligned.<br><br>**Sample Focus Area Considerations:** Create specific communication channels for rapidly sharing and escalating AI-driven tools and human analysts during real-time cybersecurity incidents.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 50; ATLAS AML.M0023 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Ensure vulnerability management is communicated and implemented throughout the organization, create communication channels to address and escalate AI-enabled attacks.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **GV.RM-06:** A standardized method for calculating, documenting, categorizing, and prioritizing cybersecurity risks is established and communicated | **General Considerations**: As threats and capabilities evolve, modulate the organization's cybersecurity risk tolerance to match the current landscape.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-18; PM-28; PM-30; RA-03 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI systems are generally more unpredictable than other types of software. The methods of calculating, categorizing, and prioritizing AI risk is treated differently from how we handle other types of software due to the scale and scope of unpredictable behavior. | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Integrate detection and control mechanisms that identify and respond to AI-specific threats (e.g. model abuse, data leakage, emerging risks), ensuring defensive systems adapt to unknown or evolving attack patterns. | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Insights regarding threats inform risk calculations and prioritization criteria.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | **Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Example Informative References:** DASF 38; ATLAS AML.M0005; ENISA Threat Landscape 2025; NIST AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | |
| **GV.RM-07:** Strategic opportunities (i.e., positive risks) are characterized and are included in organizational cybersecurity risk discussions | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-18; PM-28; PM-30; RA-03 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>Account for positive risks when discussing the use of AI in cyber defense activities.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **Roles, Responsibilities, and Authorities (GV.RR)** | Cybersecurity roles, responsibilities, and authorities to foster accountability, performance assessment, and continuous improvement are established and communicated | | | |
| **GV.RR-01:** Organizational leadership is responsible and accountable for cybersecurity risk and fosters a culture that is risk-aware, ethical, and continually improving | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-02; PM-19; PM-23; PM-24; PM-29 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 22; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 | **Proposed Priority:** 1<br><br>**Sample Opportunities:**<br><br>AI translates complex technical data into clear business language by summarizing technical risks into concise insights and helping executives quickly make informed risk decisions.<br><br>**Sample Focus Area Considerations:** Cybersecurity | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Maintain awareness of the evolving nature of AI risks and communicate new risks throughout the organization.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | Supply Chain; ENISA Threat Landscape 2025 | leaders evaluate how AI capabilities can support the needs of the cybersecurity program and ensure any new risks introduced by AI are incorporated into cybersecurity and enterprise risk processes.<br><br>Identify members of organizational leadership who approve and oversee AI-driven defense actions and policies. Review generated risk language for accuracy and alignment with organizational needs.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025 | |
| **GV.RR-02:** Roles, responsibilities, and authorities related to cybersecurity risk management are established, communicated, understood, and enforced | **General Considerations**: A human is assigned responsibility for the actions of an AI system.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-02; PM-13; PM-19; PM-23; PM-24; PM-29 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Organizations decide who is accountable for actions taken by autonomous systems.<br><br>**Example Informative References:** DASF 19, 22, 36, 41; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI agents may augment cybersecurity personnel and reduce human workload by monitoring networks, validating defensive actions, and improving detection accuracy.<br><br>**Sample Focus Area Considerations:** Define and assign responsibilities for AI-driven defense actions (e.g., automated blocking).<br><br>**Example Informative References:** | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Organizations assign personnel to roles and responsibilities for bolstering the defenses and resilience of their systems.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | DASF 39,50; ATLAS AML.M0019; ATLAS AML.M0003; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | |
| **GV.RR-03:** Adequate resources are allocated commensurate with the cybersecurity risk strategy, roles, responsibilities, and policies | **General Considerations**: Personnel have sufficient authorities and resources (e.g., budget) to meet the needs of the organization for managing its AI systems Adequate resources are allocated for personnel training on the proper use and defense against AI systems (See also PR.AT-01 and -02).<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-03 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. | **Proposed Priority:** 2<br><br>**Sample Opportunities:** Resources are allocated to support personnel with productivity, efficiency, and other enhancements (e.g., AI helps analysts to solve problems).<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Resources are available to enable layered approaches to bolster resilience and withstand attacks.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **GV.RR-04:** Cybersecurity is included in human resources practices | **General Considerations**: Personnel roles, responsibilities, and skills for managing AI systems are well defined and documented so that personnel understand the capabilities, limitations, risks, opportunities, impacts, and threats posed by AI. This information should be incorporated into hiring processes, personnel management, and role-based training where appropriate (see also PR.AT-02).<br><br>**Example Informative References:** | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Understanding the limitations of an AI model is critical to effective usage. Learn how it makes decisions, when it is likely to make mistakes, and communicate to users what to do when mistakes occur.<br><br>**Example Informative References:** DASF 33 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:**<br><br>Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/html/2511.22189v1 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** The speed at which the threat landscape may evolve is increased with AI-enabled attacks and is likely to require additional resources to help teams keep pace with staffing needs and training.<br><br>**Example Informative References:** https://arxiv.org/pdf/2305.06972 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | NIST SP 800-53, Rev 5: PM-13; PS-01; PS-07; PS-09 | | | |
| **Policy (GV.PO)** | Organizational cybersecurity policy is established, communicated, and enforced | | | |
| **GV.PO-01:** Policy for managing cybersecurity risks is established based on organizational context, cybersecurity strategy, and priorities and is communicated and enforced | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-1; SA-01; SC-01; SI-01; SR-01 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI analyzes regulatory requirements, simplifies audits and the audit process, and identifies exactly where current defenses meet regulatory standards.<br><br>**Sample Focus Area Considerations:** Incorporate rules for AI-enabled cyber defense actions including the balance of using guardrails, transparency, and HITL validation with human review and the importance of human review to combat false negatives and false positives before risk decisions are made.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **GV.PO-02:** Policy for managing cybersecurity risks is reviewed, updated, communicated, and enforced to reflect changes in requirements, threats, technology, and organizational mission | **General Considerations**: Update AI-related policy on a more frequent basis.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE- | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Policies that relate to or impact AI systems and use may require more frequent updates due to rapid changes in requirements, threats, and technological capabilities. | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI acts as a governance check by summarizing policy requirements (e.g., checking and flagging any defensive actions that conflict with requirements and rules). | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | 01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01 | **Example Informative References:** OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Sample Focus Area Considerations:** Review policies for managing cybersecurity risk with increased frequency, as AI threats are rapidly changing.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; OWASP AI Exchange: General Governance Controls | References pending additional inputs. |
| **Oversight (GV.OV)** | Results of organization-wide cybersecurity risk management activities and performance are used to inform, improve, and adjust the risk management strategy | | | |
| **GV.OV-01:** Cybersecurity risk management strategy outcomes are reviewed to inform and adjust strategy and direction | **General Considerations**: Monitor adverse events and defense mitigations to determine viability and effectiveness of defense systems.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; PM-09; PM-18; PM-30; PM-31; RA-07; SR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI acts as a governance check by summarizing policy requirements, checking and flagging any defensive actions that conflicts with rules.<br><br>**Sample Focus Area Considerations:** Regularly review AI driven defense outcomes to determine whether they reduce FPs without introducing new risks such as miss-detection<br><br>**Example Informative References:** | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Adjust use as needed to ensure detection, response, and recovery management from an AI-enabled incident is viable.<br><br>**Example Informative References:** NIST AI 100-2e2025; https://arxiv.org/html/2503.11917v3<br><br>AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | DASF 50; NIST AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | |
| **GV.OV-02:** The cybersecurity risk management strategy is reviewed and adjusted to ensure coverage of organizational requirements and risks | **General Considerations**: Update management strategy regularly as risks and opportunities present themselves.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-19; PM-30; PM-31; RA-07; SR-06 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI may require more frequent updates to the cybersecurity risk management strategy to address the rapid changes to risks and opportunities.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Updates may be required at an increased frequency to account for new attack patterns by adversarial ML research findings and results from internal read-team testing.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; OWASP AI Exchange: General Governance Controls | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Emerging attack patterns from AI-enabled tactics and techniques may necessitate new strategies and tighter feedback to keep up with rapidly changing technological capabilities.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **GV.OV-03:** Organizational cybersecurity risk management performance is evaluated and reviewed for adjustments needed | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-04; PM-06; RA-07; SR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI supports reporting by drafting clear incident summaries and compliance reports, organizing evidence, and producing standardized documentation to recuse workload of analysts. AI-generated reports should undergo human review (e.g., by analysts) for content prior to making any changes to organizational risk management.<br><br>**Sample Focus Area Considerations:** Evaluate and | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | review for adjustments needed by measuring precision, recall, and human overrides rates.<br><br>**Example Informative References:** DASF 50; OWASP AI Exchange: Model Inversion and Membership Inference; OWASP AI Exchange: Obscure Confidence; NIST AI 100-2e2025 | |
| **Cybersecurity Supply Chain Risk Management (GV.SC)** | Cyber supply chain risk management processes are identified, established, managed, monitored, and improved by organizational stakeholders | | | |
| **GV.SC-01:** A cybersecurity supply chain risk management program, strategy, objectives, policies, and processes are established and agreed to by organizational stakeholders | **General Considerations**: Organizations need to understand the origins of AI components (e.g., microservices, containers, libraries, data, hardware software), the new vulnerabilities they may introduce, and their potential impacts to cybersecurity.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-30; SR-02; SR-03; Reflections from the First Cyber AI Profile Workshop | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** With AI, data provenance should be weighted just as heavily as software and hardware origin.<br><br>All data input (both training and inference input) is an aspect of the supply chain for AI. With RL, data comes from the environment it is trained in. Take special care that the environmental conditions of the model are aligned with supply chain risk.<br><br>**Example Informative References:** DASF 22, 23, 32, 45, 51, 53; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** The integrity of training and input data (part of the data supply chain) should be verified to detect and prevent data poisoning and tempering.<br><br>In some organizations, RL can be used to detect malicious activity on a network. For example, a model that uses online learning is constantly improving from day-to-day activity. A malicious actor could work out how to trigger the alarm system in the model, while keeping the network intact and operating normally. Overtime, the model may “learn” that this behavior is creating false alarms and cease to alert operators when | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Suppliers and third parties with access to internal data, systems, and software may be the target of AI-enabled cyber attacks.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1; AI 100-2e2025 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | this behavior happens in the future. Protecting against this requires careful consideration of how the model is trained (i.e., do not allow online training or validate datasets).<br><br>**Example Informative References:** ENISA Threat Landscape 2025; NIST AI 100-2e2025 | |
| **GV.SC-02:** Cybersecurity roles and responsibilities for suppliers, customers, and partners are established, communicated, and coordinated internally and externally | **General Considerations:** Use of AI Software Bill of Materials (AI SBOM) to enhance transparency and accountability in AI components.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SR-02; SR-03; SR-05; Reflections from the First Cyber AI Profile Workshop | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 32; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1; https://arxiv.org/html/2503.11917v3 |
| **GV.SC-03:** Cybersecurity supply chain risk management is integrated into cybersecurity and enterprise risk management, risk assessment, and improvement processes | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** While the cybersecurity practices for managing supply chain risks remain the same when considering AI software, consider also that data plays a large role in the operation of AI systems and | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Integrate AI-related risk considerations in cybersecurity supply chain risk management to prevent compromised models and poisoned datasets from weakening defense capabilities. | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | PM-09; PM-18; PM-30; PM-31; SR-02; SR-03; RA-03; RA-07 | should be treated with the same gravity as software supply chains.<br><br>**Example Informative References:** DASF 13, 22, 23, 45, 51, 53; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | **Example Informative References:** ENISA Threat Landscape 2025; DASF 38,50; ENISA Threat Landscape 2025; AI 100-2e2025 | |
| **GV.SC-04:** Suppliers are known and prioritized by criticality | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: RA-09; SA-09; SR-06 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI-based services create a higher reliance on suppliers for data, compute infrastructure, software, models, or inference endpoints. This reliance increases the importance of knowing suppliers. However, the methods of identifying and prioritizing suppliers remain the same.<br><br>**Example Informative References:** DASF 13, 23, 32, 45; ATLAS AML.M0023; OWASP AI Exchange: | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Rank AI models used for cybersecurity defenses according to potential adverse impacts if compromised so defense actions can be effectively prioritized.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | | |
| **GV.SC-05:** Requirements to address cybersecurity risks in supply chains are established, prioritized, and integrated into contracts and other types of agreements with suppliers and other relevant third parties | **General Considerations**: Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SA-04; SA-09; SR-03; SR-05; SR-06; SR-10 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI systems have a larger supply chain than most software systems (e.g., more reliance on third party hardware and compute infrastructure). While it is critical to establish these requirements, the method remains the same as any other type of software.<br><br>For instance, AI systems typically rely on their training data in a more meaningful way than other software. As a result, organizations consider the supply chain for not only software and hardware, but also data.<br><br>**Example Informative References:** DASF 19, 22, 41, 51, 53; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Require vendors to disclose both model and data scope, as well as incident support for AI tools.<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **GV.SC-06:** Planning and due diligence are performed to reduce risks before entering into formal supplier or other third-party relationships | **General Considerations**: Ensure that third-party suppliers are performing their own due diligence when creating these models. This includes all the considerations that would be weighed if the model was being created in-house. Ensure that the model training methods and input data the supplier is using is accessible and transparent.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SA-04; SA-09; SR-05; SR-06 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Consider the trustworthiness of the supplier, the transparency of their data, and their training and evaluation methods before entering into third party relationships.<br><br>**Example Informative References:** DASF 22, 51, 53; DASF 51; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Before adopting a third-party AI solution, conduct AI-specific due diligence and adversarial testing to assess model vulnerabilities, ethical alignment, and performance impact.<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |
| **GV.SC-07:** The risks posed by a supplier, their products and services, and other third parties are understood, recorded, prioritized, assessed, responded to, and monitored over the course of the relationship | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: RA-09; SA-04; SA-09; SR-03; SR-06 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) While the method of handling these risks remains largely the same, regardless of whether a system is AI-enabled or not, they remain extremely important points to consider.<br><br>**Example Informative References:** DASF 7, 12, 14, 19, 29, 32, 35-39; DASF 41-42, 46, 52; 55; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Implement continuous monitoring and threat detection across supplier-provided AI models, datasets, and APIs to identify adversarial behaviors, data leakage, or compromised components originating from the supplier.<br><br>**Example Informative References:** DASF 38,39,50; ENISA Threat Landscape 2025; AI 100-2e2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | | |
| **GV.SC-08:** Relevant suppliers and other third parties are included in incident planning, response, and recovery activities | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SA-04; SA-09; SR-02; SR-03; SR-08; CP-01; IR-01 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) While the method of supplier collaboration remains the same in an AI-context, the importance of establishing these relationships remains high.<br><br>**Example Informative References:** DASF 23, 39, 45; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI collects, formats, and shares threat intelligence using standards protocols (such as STIX and OpenCTI) to ensure teams and partners stay aligned.<br><br>**Sample Focus Area Considerations:** Integrate suppliers and partners, particularly those providing AI models or data, into AI-specific incident response and recovery plans, focusing on coordinated detection of adversarial attacks and data poisoning.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) As AI is integrated into defensive practices through third-party suppliers to bolster cyber defenses, such as those for endpoint detection services, it is critical for organizations to include suppliers in incident planning and response.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |
| **GV.SC-09:** Supply chain security practices are integrated into cybersecurity and enterprise risk management programs, and their performance is monitored throughout the technology product and service life cycle | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-19; PM-28; PM-30; PM-31; RA-03; RA-07; SA-04; SA-09; SR-02; SR-03; SR-05; SR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 23, 45, 51, 53; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 38,50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

| CSF 2.0 Core: GOVERN | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | Supply Chain; ENISA Threat Landscape 2025 | | |
| **GV.SC-10:** Cybersecurity supply chain risk management plans include provisions for activities that occur after the conclusion of a partnership or service agreement | **General Considerations**: Review and understand terms of service for AI considerations (e.g., data ownership, permissible uses data) before using third-party resources.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-31; RA-03; RA-05; RA-07; SA-04; SA-09; SR-02; SR-03; SR-05; SR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 13, 51, 53; ATLAS AML.M0023; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-161 Rev. 1 |

## 2.4. Cyber AI Profile: IDENTIFY

**Table 2 Cyber AI Profile – IDENTIFY.**

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **IDENTIFY (ID)** | The organization's current cybersecurity risks are understood | | | |
| **Asset Management (ID.AM)** | Assets (e.g., data, hardware, software, systems, facilities, services, people) that enable the organization to achieve business purposes are identified and managed consistent with their relative importance to organizational objectives and the organization's risk strategy | | | |
| **ID.AM-01:** Inventories of hardware managed by the organization are maintained | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-08; PM-05 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 10-11, 18-19, 29, 34; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Inventory the accelerated computes used to conduct AI-enabled cyber defense. Tracking these assets helps ensure sufficient compute capacity for defense, operations and support faster scoping and containment during incident response.<br><br>**Example Informative References:** DASF 1,34; ATLAS AML.M0023; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:**<br><br>NIST SP 800-172 Configuration Management 3.4.1e-3.4.3e; |
| **ID.AM-02:** Inventories of software, services, and systems managed by the organization are maintained | **General Considerations**: AI capabilities are increasingly embedded into existing systems, which may introduce challenges to maintaining a current inventory.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-20; CM-08; PM-05; SA-05; SA-09 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Identifying systems that have embedded AI capabilities may not always be straightforward, especially those purchased by a third party.<br><br>**Example Informative References:** DASF 10-11, 18-19, 23, 29-30, 32, 34, 45, 48; OWASP AI Exchange: | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Inventories should include AI models, APIs, keys, agents, data (see also ID.AM-07), and their integrations and permissions.<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025;AI 100-2e2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:**<br><br>NTIA SBOM (all); NIST SP 800-172 Configuration Management 3.4.1e-3.4.3e; NIST SP 800-218 (all); ATT&CK M1047 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | | |
| **ID.AM-03:** Representations of the organization's authorized network communication and internal and external network data flows are maintained | **General Considerations:** No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-04; CA-03; CA-09; PL-02; PL-08; PM-07 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Categorize traffic into four distinct groups: internal human generated traffic, internal computer-generated network traffic (such as from cron jobs or automated processes), internal AI based traffic (AI tools searching the web or utilizing organizational resources), and external traffic.<br><br>Note that external traffic is difficult to categorize as human or bot generated. However, there is value in tracking network traffic surrounding model registries and dataset sources, to better detect attempted supply chain attacks.<br><br>**Example Informative References:** DASF 5, 7, 10-11, 13, 16, 18-19, 24, 28, 30-32, 41, 44, 48, 52, 56, 58-60, 62; OWASP AI Exchange: | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Maintain representations of AI data flows, including the paths for inference requests, training data pipelines to enforce defense boundaries and detect anomalies.<br><br>**Example Informative References:** DASF 38; ENISA Threat Landscape 2025; AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. Understanding the network facilitates a clearer understanding of normal and abnormal traffic that can indicate a security event is occurring.<br><br>**Example Informative References:**<br><br>NIST SP 800-172 Configuration Management 3.4.1e-3.4.3e; ATT&CK M1015; ATT&CK M1028; AI 100-2e2025 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | | |
| **ID.AM-04:** Inventories of services provided by suppliers are maintained | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-20; SA-09; SR-02 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) While it is important to inventory third party services, the methods do not change in an AI-context.<br><br>**Example Informative References:** DASF 2, 5, 18, 23-24, 29, 32-33, 42, 45, 48, 64; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** External AI defense components and services (e.g., detection models) provided by suppliers should be listed in service inventories.<br><br>**Example Informative References:** DASF 45,48; ATLAS AML.M0005; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. Up-to-date inventories enable tracking and management of services provided by suppliers and third parties who may be affected by, or the cause of, AI-enabled attacks.<br><br>**Example Informative References:** ATLAS AML.M0023; NTIA SBOM (all); NIST SP 800-218 (all); ATT&CK M1033; AI 100-2e2025 |
| **ID.AM-05:** Assets are prioritized based on classification, criticality, | **General Considerations**: No general considerations identified - see Focus Area Considerations. | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** New | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Prioritize AI assets | **Proposed Priority:** 3<br>**Sample Focus Area** |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| resources, and impact on the mission | **Example Informative References:** NIST SP 800-53, Rev 5: RA-03; RA-09; RA-02 | classifications will be needed for AI systems. These systems also consume and use resources which directly contribute to the functionality and vulnerability of the system (such as the training dataset), in a way that other software systems are not typically impacted.<br><br>**Example Informative References:** DASF 5-6, 23-24, 30, 45; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | (e.g. Models, inference services) based on their criticality, the mission impact, and their data classification for AI-enabled defense.<br><br>**Example Informative References:** DASF 38; ENISA Threat Landscape 2025 | **Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **ID.AM-07:** Inventories of data and corresponding metadata for designated data types are maintained | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-12; CM-13; SI-12 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Data provenance and data inventories play a mutually supportive role in understanding the nature of the data and metadata, as well as the requirements that travel with them (e.g. use agreements, consent). This can necessitate additional diligence for data that is collected from external sources or that will be shared (either externally or even internally for different purposes). The NIST Privacy Framework can be applied to help organizations manage privacy risk. | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI technologies enhance capabilities that can help organizations more rapidly understand the data and metadata they are managing, including where it resides on the network, characterize the nature of it, and align it with compliance and other risk management requirements (e.g., identifying data that is subject to protections under a law or regulation).<br><br>**Sample Focus Area Considerations:** Understanding the data and their corresponding meta data (e.g., their location, protection needed) is important | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI 100-2e2025 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | New metadata will need to be tracked to capture the full picture of data used for ML. For instance, what in-memory data transformations/augmentations are taking place to train a model? Maintain provenance of AI datasets. Modified data could indicate data poisoning has occurred. Training data originates outside the organization can introduce additional uncertainties.<br><br>**Example Informative References:** DASF 5-6, 10-11, 15, 17, 21, 30, 32, 48, 56, 58-59, 62; ENISA Threat Landscape 2025; AI 100-2e2025 | for effective AI-enable defense. AI can help to discover data automatically, label and classify, and identify information that needs special handling.<br><br>**Example Informative References:** DASF 5,38; ENISA Threat Landscape 2025; AI 100-2e2025 | |
| **ID.AM-08:** Systems, hardware, software, services, and data are managed throughout their life cycles | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-09; CM-13; MA-02; MA-06; PL-02; PM-22; PM-23; SA-03; SA-04; SA-08; SA-22; SI-12; SI-18; SR-05; SR-12 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Data plays an especially important role in the performance of AI-based systems. Particularly for systems that may be making automated decisions or contributing data to other processes (e.g., training data), ensuring data quality throughout the lifecycle of that data and system can be especially important to ensure mistakes and “bad” data are not proliferated. | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 6,38,50; ENISA Threat Landscape 2025; AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Maintaining systems throughout their lifecycle will promote resilience and ensure effectiveness when/if AI-enabled attacks try to leverage outdated or mismanaged components for new attack surfaces.<br><br>**Example Informative References:** NIST SP 800-281 (all); NTIA SBOM |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | **Example Informative References:** DASF 6-7, 10-12, 14, 16-20, 22-23, 29-30, 32, 34, 38, 41-42, 48, 52, 59, 63; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP Model Input Confidentiality; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | | (all); ATLAS AML.M0023; ATT&CK M1054; AI 100-2e2025 |
| **Risk Assessment (ID.RA)** | The cybersecurity risk to the organization, assets, and individuals is understood by the organization | | | |
| **ID.RA-01:** Vulnerabilities in assets are identified, validated, and recorded | **General Considerations**: AI introduces new classes of vulnerabilities (i.e., adversarial input) which organizations account for when securing AI applications and using AI systems to defend organizations.<br><br>AI systems can also leverage existing vulnerabilities in software.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CA-02; CA-07; CA-08; RA-03; RA-05; SA-11(02); SA-15(07); SA-15(08); SI-04; SI-05 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** New classes of vulnerabilities exist with AI, such as adversarial input. This introduces a new suite of vulnerabilities to be identified, tracked, and recorded.<br><br>**Example Informative References:** DASF 13, 16, 19, 22-23, 31-32, 36, 38, 41-42, 45-46, 52-53, 56, 63; OWASP Conventional Runtime Controls; OWASP General; https://arxiv.org/pdf/2409.08831v1 Governance; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2504. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Include AI-specific attacks (e.g., adversarial input, model evasion) in vulnerability management and flag AI-enabled attacks documented in recent threat reports to ensure defenses are current.<br><br>**Example Informative References:** DASF 38,39; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029; https://arxiv.org/pdf/2409.08831v1; AI 100-2e2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Vulnerabilities in assets such as software can potentially be exploited by AI-enabled attacks, necessitating a faster identification and resolution time than previous cybersecurity policy.<br><br>**Example Informative References:** NIST SP 800-218 (all); NTIA SBOM (all); ATLAS AML.M0016; ATT&CK M1016; ATT&CK M1051; https://arxiv.org/pdf/2504.11168; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | 11168; https://arxiv.org/pdf/2309.01029; AI 100-2e2025 | | |
| **ID.RA-02:** Cyber threat intelligence is received from information sharing forums and sources | **General Considerations**: Incorporate new information sharing sources related to AI, such as AI-specific threat intelligence from publications, journals, ISACs/ISAOs,[6] and other forums. Examples of available resources include: OWASP,[7] The AI Incident Database (AIID), and MITRE ATLAS™.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SI-05; PM-15; PM-16 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Some information which could have immediate real-world impact on models may come from AI academic publications. For instance, AI jailbreaking and prompt injection techniques are more likely to be discovered and published in AI-related journals when compared to other sources of cybersecurity information.<br><br>**Example Informative References:** ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI helps defenders by scanning massive datasets, rapidly analyzing large volumes of threat intelligence, and correlating the information from multiple sources. This gives defenders a clear and more predictive view of the threat landscape.<br><br>**Sample Focus Area Considerations:** Retrieve AI-focused CTI, such as jailbreaks and prompt injection, from research and industry-shared resources (e.g., ATLAS™).<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Information from red-team and penetration testing exercises as well as AI-related academic publications and should be used to inform future capabilities for thwarting AI-enabled attacks.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |
| **ID.RA-03:** Internal and external threats to the organization are identified and recorded | **General Considerations**: AI presents a myriad of new internal and external threats to an organization, but also opportunities to defend against these threats. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** LLM-based systems can operate autonomously and are sometimes given the ability to | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Due to the increased risk of AI-enabled phishing and social engineering tactics and techniques, emphasize |

[6] America's AI Action Plan, released in July 2025, includes a recommendation for the U.S. Department of Homeland Security lead establishment of an AI-ISAC in collaboration with NIST's Center for AI Standards and Innovation (CAISI) (under the U.S. Department of Commerce), and the Office of the National Cyber Director.
[7] For example, OWASP makes "top 10" lists available for LLM and ML.

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: PM-12; PM-16; RA-03; SI-05 | execute arbitrary code inside of a compute environment. The potential uncertainties around autonomous behaviors (e.g., being manipulated into code execution) should be considered alongside other threats.<br><br>AI systems are not perfectly predictable. These systems may have unexpected output when presented with input outside their training domain.<br><br>**Example Informative References:** DASF 13, 32; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | unusual patterns faster than other tools.<br><br>AI helps defenders by scanning massive datasets and rapidly analyzing large volumes of threat intelligence to identify internal and external threats across IT, OT, and IOT, giving defenders a clear view of the threat landscape.<br><br>**Sample Focus Area Considerations:** Account for internal and external AI-enabled threats that can directly target or mislead defensive systems, such as AI-enabled phishing, deepfakes, and agent manipulation. Defensive teams should establish processes to identify, log, and analyze these threats.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 39; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | awareness and training that is up-to-date and encompasses new and emerging trends that exploit personnel via email, chatbots, and audio/visual content generated by AI.<br><br>**Example Informative References:** ATLAS; ATT&CK; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **ID.RA-04:** Potential impacts and likelihoods of threats exploiting vulnerabilities are identified and recorded | **General Considerations**: AI introduces new distinct considerations across all Focus Areas. Resources like and MITRE ATLAS™ can help organizations understand their impacts. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI systems present many new attack vectors such as adversarial input, data leakage, data poisoning, error- | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Include new risk vectors that emerge with the use of AI-enabled cyber defense (e.g., | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI-enabled attacks will target existing vulnerabilities and aid in vulnerability discovery—risk |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-11; RA-02; RA-03; RA-08; RA-09 | amplifying feedback loops, and concept drift.<br><br>**Example Informative References:** DASF 6, 13, 22-23, 45; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2504.11168; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | data leakage, overreliance, and excessive agency) in risk modeling.<br><br>**Example Informative References:** DASF 38,39;OWASP AI Exchange: How about Generative AI?; ATLAS AML.M0005; https://arxiv.org/pdf/2309.01029; AI 100-2e2025 | analysis and vulnerability management should be prioritized.<br><br>**Example Informative References:** NIST SP 800-218 (all); ATLAS AML.M0016; ATT&CK M1051; https://arxiv.org/pdf/2309.01029; AI 100-2e2025 |
| **ID.RA-05:** Threats, vulnerabilities, likelihoods, and impacts are used to understand inherent risk and inform risk response prioritization | **General Considerations:** Prioritize controls based on threats, likelihood, and impact to reduce high risk AI errors.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-16; RA-02; RA-03; RA-07 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 6, 13, 19, 36; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029 ; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI translates complex technical data into clear business language by summarizing technical risks into concise insights and helping executives make informed risk decisions.<br><br>**Sample Focus Area Considerations:** Continuously evaluate AI cybersecurity defense capabilities to ensure they are sufficiently mature to use for their intended purpose before deployment.[8]<br><br>**Example Informative References:** | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Incorporate the impact of AI-enabled attacks when prioritizing risk responses.<br><br>**Example Informative References:** https://arxiv.org/pdf/2309.01029 ; https://arxiv.org/html/2503.11917v3 |

[8] One example that illustrates technology readiness is: https://cloud.google.com/blog/products/identity-security/rsa-introducing-ai-powered-insights-threat-intelligence

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | DASF 38; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2309.01029; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | |
| **ID.RA-06:** Risk responses are chosen, prioritized, planned, tracked, and communicated | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-18; PM-30; RA-07 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 12-13, 36, 38; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority: 2**<br><br>**Sample Opportunities:** AI can be used for proactive risk management and predicting incident outcomes which can enable faster responses, inform risk responses, and support personnel with communicating those decisions. AI analyzes vulnerabilities to issue early warnings and helps prioritize defensive actions before incidents occur.<br><br>**Sample Focus Area Considerations:** Define conditions for disabling AI autonomy during risk response, while also AI predictive incident outcomes to prioritize and communicate risk responses.<br><br>**Example Informative References:** DASF 38,39; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) The baseline risks and threats that organizations face is increased by the existence of AI, and AI also lowers the barriers to entry for carrying out cyber attacks. However, the cybersecurity controls that an organization should implement for this Subcategory remain the same.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |
| **ID.RA-07**: Changes and exceptions are managed, | **General Considerations**: It is difficult to predict how a change | **Proposed Priority:** 2 | **Proposed Priority:** 1 | **Proposed Priority:** 1 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| assessed for risk impact, recorded, and tracked | to an AI system will impact the overall functionality. Organizations should use version control processes to record any change to the hardware, software, configuration values, or training data. Review and log changes to the model, training data, and the model configuration settings (e.g., hyperparameters) for cybersecurity defenses that use AI.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CA-07; CM-03; CM-04 | **Sample Focus Area Considerations:** Even small changes to AI systems may introduce risks which are unknowable ahead of time. It can be difficult (or impossible) to predict the impacts of changes to a model without making the change and testing the resultant AI.<br><br>**Example Informative References:** DASF 7, 13-18, 29, 41, 52; OWASP AI Exchange: General Governance Controls; ENISA Threat Landscape 2025 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Cybersecurity defenses play a critical role in protecting the organization.<br><br>**Example Informative References:** DASF 50; ENISA Threat Landscape 2025; AI 100-2e2025 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Changes, even when minor, could introduce weaknesses into the system that are exploitable by AI-enabled attacks.<br><br>**Example Informative References:** NIST SP 800-172 - Configuration Management (3.4.1e-3.4.3e); NIST SP 800-172 – Security Assessment (3.12.1e); ATT&CK M1033; ATT&CK M1054 |
| **ID.RA-08:** Processes for receiving, analyzing, and responding to vulnerability disclosures are established | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: RA-05 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 12-13, 19, 53; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 38,50; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Vulnerability management, disclosures, and response will need higher prioritization due to speed and efficiency of AI-enabled attacks to exploit vulnerabilities at scale.<br><br>**Example Informative References:** NIST SP 800-218 (all); ATT&CK M1016; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **ID.RA-09:** The authenticity and integrity | **General Considerations:** No general considerations | **Proposed Priority:** 2 | **Proposed Priority:** 2 | **Proposed Priority:** 2 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| of hardware and software are assessed prior to acquisition and use | identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SA-04; SA-05; SA-10; SA-11; SA-15; SA-17; SI-07; SR-05; SR-06; SR-10; SR-11 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) While the methods of verifying authenticity and integrity do not change, they take on elevated importance in AI-contexts due to the lack of visibility into third-party tools.<br><br>**Example Informative References:** DASF 23, 45, 52; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP Model input confidentiality; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 50; OWASP AI Exchange: General Governance Controls; ENISA Threat Landscape 2025 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Extra measures may be necessary to prevent use of counterfeit hardware and software to avoid using unapproved or malicious AI.<br><br>**Example Informative References:** NIST SP 800-218 (all); ATT&CK M1033; ATT&CK M1054; ATT&CK M1044; AI 100-2e2025 |
| **ID.RA-10:** Critical suppliers are assessed prior to acquisition | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12-14, 19, 29, 35-39, 41-42, 46, 51-53, 55; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 50; OWASP AI Exchange: General Governance Controls | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Suppliers and third parties with access to internal data, systems, and software may be the target of AI-enabled cyber attacks.<br><br>**Example Informative References:** NIST SP 800-218 (all); ATT&CK M1047 |
| **Improvement (ID.IM)** | Improvements to organizational cybersecurity risk management processes, procedures and activities are identified across all CSF Functions | | | |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **ID.IM-01:** Improvements are identified from evaluations | **General Considerations**: AI introduces a suite of new metrics to consider. Carefully select which metrics closely align with the organization's goals. For instance, perhaps FPs are much more important than false negatives, in which case evaluation metrics will be more heavily influenced by precision rather than recall, and using metrics such as an F1 score may obfuscate results.<br><br>Incorporate feedback loops to continue improving AI use in cybersecurity activities.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; CA-02; CA-05; CA-07; CA-08; CP-02; IR-04; IR-08; PL-02; RA-03; RA-05; RA-07; SA-08; SA-11; SA-17(06); SI-02; SI-04; SR-05 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Evaluations in ML need to be carefully weighed and considered. Consider the structure and form of the evaluation datasets, as well as what biases may exist in that data. Also consider identifying appropriate metrics to measure a model's performance. Oftentimes, accuracy does not fully reflect adequacy of performance.<br><br>**Example Informative References:** OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP Model Input Controls; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI supports reporting by drafting clear incident summaries and compliance reports, organizing evidence, and producing standardized documentation to reduce workload of analysts.<br><br>**Sample Focus Area Considerations:** Identify continuous security improvements by evaluating AI model performance metrics (e.g., precision/recall, drift rates) against security goals, such as FP reduction and adversarial robustness.<br><br>**Example Informative References:** DASF 38,39,50; OWASP AI Exchange: Model Access Control; ATLAS AML.M0008; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Organizations may consider leveraging and implementing AI-assisted defenses to maintain pace and scale of AI-enabled cyber attacks. See the opportunities and considerations of the Defend Focus Area for guidance.<br><br>**Example Informative References:** https://ieeexplore.ieee.org/document/10747338; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **ID.IM-02:** Improvements are identified from security tests and exercises, including those | **General Considerations:** Improvements are identified from AI Red Teaming exercises and through coordinated testing with | **Proposed Priority:** 3 | **Proposed Priority:** 3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| done in coordination with suppliers and relevant third parties | suppliers of AI models and data to harden the supply chain.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; CA-02; CA-05; CA-07; CA-08; CP-02; CP-04; IR-03; IR-04; IR-08; PL-02; PM-04; PM-31; RA-03; RA-05; RA-07; SA-08; SA-11; SI-02; SI-04; SR-05 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 13; DASF 16; DASF 19; DASF 36; DASF 39; DASF 41; DASF 44; DASF 46; DASF 56; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP Model Input Controls; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39; OWASP AI Exchange: Model Access Control; ATLAS AML.M0018; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | cybersecurity practices apply. (Rationale) Organizations may consider leveraging and implementing AI-assisted penetration-testing and red-teaming tools to maintain pace and scale of AI-enabled cyber-attacks when performing security tests.<br><br>**Example Informative References:** NIST SP 800-172 – Security Assessment (3.12.1e); https://arxiv.org/html/2503.11917v3 |
| **ID.IM-03:** Improvements are identified from execution of operational processes, procedures, and activities | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AT-01; AU-01; CA-01; CM-01; CP-01; IA-01; IR-01; MA-01; MP-01; PE-01; PL-01; PM-01; PS-01; PT-01; RA-01; SA-01; SC-01; SI-01; SR-01; CA-02; CA-05; CA-07; CA-08; CP-02; IR-04; IR-08; PL-02; PM-04; | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 5; DASF 13; DASF 16; DASF 18; DASF 19; DASF 24; DASF 29; DASF 33; DASF 36; DASF 38; DASF 39; DASF 41; DASF 42; DASF 44; DASF 46; DASF 51; DASF 55; DASF 56; DASF 60; DASF 64; OWASP AI Exchange: Controls to Limit the | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Operational feedback, such as high human override rates, frequent model drift alerts, or patterns of adversarial input detection, should be analyzed to drive continuous security improvements in AI-enabled defense systems.<br><br>**Example Informative References:** | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | PM-31; RA-03; RA-05; RA-07; SA-04; SA-08; SA-11; SI-02; SI-04 | Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP Model Input Controls; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | DASF 21,39; ATLAS AML.M0008; ATLAS AML.M0015; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | |
| **ID.IM-04:** Incident response plans and other cybersecurity plans that affect operations are established, communicated, maintained, and improved | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-02; IR-08; PL-02; SR-02 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Incidents involving AI systems have specific response procedures including but not limited to reducing the scope of access a model has (whether that be data, tools, or networks), identifying the issue via logs, and restoring stable model backups (beyond using code versioning, but also model versioning and potentially even previous dataset versions).<br><br>**Example Informative References:** DASF 19; DASF 39; DASF 41; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Incident response plans include AI-specific procedures for containment (e.g., disabling model autonomy), triage (e.g., analyzing model logs), and recovery (e.g., restoring validated model versions).<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025; ATLAS AML.M0014; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Improvements to these plans need to reflect evolutions in AI-enabled attacks and (any) effective mitigations or actions taken. This can help improve the efforts related to PR.AT-01 and specifically PR.AT-02, as well as continuous monitoring and anomalous activity detection. Updates here are most likely more dynamic due to the speed at which AI-enabled attacks can be executed but the speed of sophistication and deception.<br><br>**Example Informative References:** NIST SP 800-172 – Incident |

| CSF 2.0 Core: IDENTIFY | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | | Response (3.6.1e-3.6.2e); AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |

## 2.5. Cyber AI Profile: PROTECT

**Table 3 Cyber AI Profile - PROTECT.**

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **PROTECT (PR)** | Safeguards to manage the organization's cybersecurity risks are used | | | |
| **Identify Management, Authentication, and Access Control (PR.AA)** | Access to physical and logical assets is limited to authorized users, services, and hardware and managed commensurate with the assessed risk of unauthorized access | | | |
| **PR.AA-01:** Identities and credentials for authorized users, services, and hardware are managed by the organization | **General Considerations**: Give AI systems unique and traceable identities and credentials to better track their activity.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AC-02; AC-14; IA-01; IA-02; IA-03; IA-04; IA-05; IA-06; IA-07; IA-08; IA-09; IA-10; IA-11 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI systems may need their own identities and credentials (i.e., AI service level accounts) to interact with a broader system. Organizations need traceability between AI systems and their actions.<br><br>**Example Informative References:** DASF 1, 3, 21-22, 32, 40, 48; ATLAS AML.M0005; ATLAS AML.M0019; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI catches credential misuse that previous rules might miss by flagging unusual authentication activity.<br><br>**Sample Focus Area Considerations:** Assign and manage unique and traceable identities and credentials to AI defense agents to support defensive response activities.<br><br>**Example Informative References:** DASF 39; WASP AI Exchange: Model Access Control; ENISA | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI-enabled cyber attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware.<br><br>**Example Informative References:** NIST SP 800-172 – Configuration Management (3.4); NIST SP 800-172 Identification and Authentication (3.5); NIST SP 800-207 – Zero Trust Architecture; |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | Threat Landscape 2025; ATLAS AML.M0005 | ATT&CK M1032; ATT&CK M1027; ATLAS AML.M0005; ATLAS AML.M0019; AI 100-2e2025 |
| **PR.AA-02:** Identities are proofed and bound to credentials based on the context of interactions | **General Considerations:** LLMs or AI Agents may have access to data sources or tools that a user typically would not have. Bind agent and service identities to their credentials using cryptographic signing and mutual authentication.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IA-12 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 1, 3-5, 22, 32-33, 40, 57; ATLAS AML.M0005; ATLAS AML.M0026; ATLAS AML.M0027; ATLAS AML.M0028; OWASP Conventional Runtime Controls; AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI services should be bound in context and time to prevent excessive agency. Prevent spoofing by requiring certificate-based validation and continuous monitoring of agent behavior so that malicious tools cannot imitate the organization's defense agents.<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025; ATLAS AML.M0013; AML.M0014); AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI-enabled cyber attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **PR.AA-03:** Users, services, and hardware are authenticated | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-07; AC-12; IA-02; IA-03; IA-05; IA-07; IA-08; IA-09; IA-10; IA-11 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard Cybersecurity practices apply.<br><br>**Example Informative References:** DASF 1-2, 4-5, 18, 21-22, 24, 29, 31-33, 40, 42-43, 46, 56, 64; ATLAS AML.M0005; OWASP | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Assign each AI agent with a unique identity and credentials and treat them with the same security precautions as privileged users.<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025; ATLAS | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI-enabled cyber-attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware, but the way that |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | Conventional Runtime Controls; ENISA Threat Landscape 2025 | AML.M0019; ATLAS AML.M0005; AI 100-2e2025 | these attacks are mitigated remains the same.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **PR.AA-04:** Identity assertions are protected, conveyed, and verified | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IA-13 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 22, 40 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Identity assertions provide a measure of assurance that AI components are valid. Signing and verifying agent assertions and tokens provide insight into their provenance and supports organizations in verifying that only the expected agents are operating.<br><br>**Example Informative References:** DASF 39; OWASP AI Exchange: Monitor Use; ATLAS AML.M0013; ATLAS AML.M0014; AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) AI-enabled cyber-attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware, but the way that these attacks are mitigated remains the same.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **PR.AA-05:** Access permissions, entitlements, and authorizations are defined in a policy, managed, enforced, and reviewed, and incorporate the principles of least privilege and separation of duties | **General Considerations**: AI systems should be treated separately from other types of entities within a network and require their own set of permissions and authorization policies. Apply the principle of least privilege to AI agents by granting only the permissions necessary to carry out their role. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Because AI systems can interact in complex and unpredictable ways, they may need a new set of policies to govern their permissions and authorizations.<br><br>**Example Informative References:** DASF 2-3, 5-6, 14, 18, 21, 22, 24, | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI catches credential misuse that previous rules might miss by flagging unusual authentication activity.<br><br>**Sample Focus Area Considerations:** Establish processes to manage AI agent privileges over time (e.g., periodic | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Maintain robust policies for access controls and authorizations to prevent AI-enabled cyber-attacks that seek to obtain credentials, access keys/tokens, etc., through |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: AC-01; AC-02; AC-03; AC-05; AC-06; AC-10; AC-16; AC-17; AC-18; AC-19; AC-24; IA-13 | 29, 31-33, 40, 42-43, 46, 51, 56-58, 60, 64; ATLAS AML.M0005; ATLAS AML.M0012; ATLAS AML.M0026; ATLAS AML.M0027; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 https://arxiv.org/pdf/2504.11168; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | reviews and updates) like other types of privileged accounts.<br><br>**Example Informative References:** OWASP AI Exchange: Least Model Privilege; DASF 39,50; ENISA Threat Landscape 2025; ATLAS AML.M0005; ATLAS AML.M0019; ATLAS AML.M0026; ATLAS AML.M0027; ATLAS AML-M0028; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | methods such as scalable brute force attacks.<br><br>**Example Informative References:** AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **PR.AA-06:** Physical access to assets is managed, monitored, and enforced commensurate with risk | **General Considerations:** No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PE-02; PE-03; PE-04; PE-05; PE-06; PE-08; PE-18; PE-19; PE-20 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:**<br><br>Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 1, 4-5, 9, 16, 21-22, 24, 31, 40, 46, 56, 57, 59, 62; ATLAS AML.M0005; ATLAS AML.M0012 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Leverage AI-enabled systems to strengthen physical defense capabilities, such as the use of computer vision to detect unauthorized access, tampering, or intrusions in data centers and Graphics Processing Unit (GPU) racks.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 39; ATLAS AML.M0005 ; ATLAS AML.M0009 ; AI 100-2e2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:**<br><br>Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **Awareness and Training (PR.AT)** | The organization's personnel are provided with cybersecurity awareness and training so that they can perform their cybersecurity-related tasks | | | |
| **PR.AT-01:** Personnel are provided with awareness and training so that they possess the knowledge and skills to perform general tasks with cybersecurity risks in mind | **General Considerations**: Personnel should be adequately trained to work with the results of AI systems, which are evolving rapidly and sometimes emit unpredictable output.<br><br>This training will need to be frequently updated and readministered to match the pace of developments with AI technology.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AT-02; AT-03 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI presents a new dimension to the cybersecurity landscape which has not been seen before. Personnel should be aware of new considerations, such as model limitations, adversarial input, and concept drift as well as how these apply to the specific context.<br><br>**Example Informative References:** DASF 25, 32, 39, 41, 61; ATLAS AML.M0012; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** The use of AI presents risks to the accuracy of information, such as hallucination and confabulation. Train analysts to monitor agents and how to evaluate AI outputs to spot hallucinations, bias, manipulated responses, and other potential issues before acting.<br><br>**Example Informative References:** DASF 25,31,32,29,41,6; https://arxiv.org/pdf/2311.05232; ENISA Threat Landscape 202; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI presents a new dimension to the cybersecurity threat landscape which has not been seen before. Personnel should be aware of and have access to training to inform them about new, emerging, AI-enabled threats such as those leveraging spear phishing and social engineering tactics and techniques.<br><br>**Example Informative References:** NIST SP 800-172 – Awareness and Training (3.2); ATT&CK M1017; ATLAS AML.M0018; NIST AI 600-1 MP-5.1-002; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **PR.AT-02:** Individuals in specialized roles are provided with awareness and training so that they possess the knowledge and skills to perform relevant tasks with cybersecurity risks in mind | **General Considerations**: Specialized roles responsible for the use of AI systems or protection against AI-enabled attacks are aware of both cybersecurity and AI-specific threats and associated mitigation strategies.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AT-03 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI presents a new dimension to the cybersecurity landscape. Personnel responsible for protecting AI systems should be aware of both cybersecurity and AI-specific considerations that | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI creates realistic attack simulations and phishing scenarios to test defenses and help defenders practice and improve response skills.<br><br>**Sample Focus Area Considerations:** Red teaming and personnel training should address | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI presents a new dimension to the cybersecurity threat landscape. Personnel should be aware of and have access to training and information on new, emerging, AI-enabled threats such as those leveraging spear phishing and social |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | are applicable to the organization's context.<br><br>**Example Informative References:** DASF 25, 39, 61; ATLAS AML.M0012; ENISA Threat Landscape 2025; AI 100-2e2025 | AI-enabled defense actions, adversarial ML, prompt injection, model drift, and AI forensics to support personnel in building skills and knowledge to use of AI-enabled defense actions that align with organizational need and risk tolerance.<br><br>**Example Informative References:** DASF 25,32,39,41,61; ATLAS AML.M0018; AML.M0003; AI 100-2e2025 | engineering tactics and techniques. Up-to-date training on AI-enabled phishing, social engineering, and use of chatbots is critical to thwart cybersecurity incidents. Personnel in specialized roles (e.g., incident response) will need additional training to recognize and validate AI-enabled attacks. Modeling and simulation scenarios may also need to be developed to facilitate effective detection, response, and recovery efforts.<br><br>**Example Informative References:** NIST SP 800-172 – Awareness and Training (3.2); ATT&CK M1017; ATLAS AML.M0018; NIST AI 600-1 MP-5.1-002; AI 100-2e2025 |
| **Data Security (PR.DS)** | Data are managed consistent with the organization's risk strategy to protect the confidentiality, integrity, and availability of information | | | |
| **PR.DS-01:** The confidentiality, integrity, and availability of data-at-rest are protected | **General Considerations**: Data is critical for the functionality of many AI systems. Examples of AI-specific confidentiality, integrity, and availability concerns include: availability poisoning and availability breakdown, integrity poisoning, and confidentiality leakage and compromises (which can also result in privacy issues). If data is corrupted, then ML models will no longer train with that data. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) See General Considerations.<br><br>**Example Informative References:** DASF 3-13, 16-17, 20-22, 24-27, 29-31, 33, 36, 42, 44, 46, 51, 56, 58-60, 62**;** ATLAS AML.M0005; ATLAS AML.M0012; ATLAS AML.M0024; ATLAS AML.M0025; | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Enabling accurate, scaled, and fast defensive operations require trusted data.<br><br>**Example Informative References:** DASF 32,39,41; ENISA Threat Landscape 2025 ATLAS AML.M0012; ATLAS AML.M0005; AI 100-2e2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) See General Considerations.<br><br>**Example Informative References:** ATT&CK M1057; ATT&CK M1053; ATT&CK M1041; ATT&CK M1029; ATLAS AML.M0007; AI 100-2e2025 |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | If the data is tampered with, then models will not train well. While these considerations are of utmost importance to AI, the mitigations an organization might use to prevent them remain the same as other types of software encryption and data backups.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CA-03; CP-09; MP-08; SC-04; SC-07; SC-12; SC-13; SC-28; SC-32; SC-39; SC-43; SI-03; SI-04; SI-07 | OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP Model Input Confidentiality; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | | |
| **PR.DS-02:** The confidentiality, integrity, and availability of data-in-transit are protected | **General Considerations**: Examples of AI-specific confidentiality, integrity, and availability concerns include: availability poisoning and availability breakdown, integrity poisoning, and confidentiality leakage and compromises (which can also result in privacy issues).<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-16; CA-03; SC-04; SC-07; SC-08; SC-11; SC-12; SC-13; SC-16; SC-40; SC-43; SI-03; SI-04; SI-07 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 3-5, 7, 9, 11, 13, 16, 21-27, 29-34, 36, 44-46, 51, 56, 58-60, 62; ATLAS AML.M0024; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP Model Input Confidentiality; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References: DASF 25,37,39;** ENISA Threat Landscape 2025; ATLAS AML.M0019 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 9; ATT&CK M1057; |
| **PR.DS-10:** The confidentiality, integrity, | **General Considerations**: AI always has the potential to leak sensitive | **Proposed Priority:** 1 | **Proposed Priority:** 1 | **Proposed Priority:** 3 |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| and availability of data-in-use are protected | data if it has access to it. Examples of AI-specific confidentiality, integrity, and availability concerns include: availability poisoning and availability breakdown, integrity poisoning, and confidentiality leakage and compromises (which can also result in privacy issues). In general, an AI model has the same level of sensitivity as the data it was trained on and the data that it has access to at inference.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-02; AC-03; AC-04; AU-09; AU-13; CA-03; CP-09; SA-08; SC-04; SC-07; SC-11; SC-13; SC-24; SC-32; SC-39; SC-40; SC-43; SI-03; SI-04; SI-07; SI-10; SI-16 | **Sample Focus Area Considerations:** The information that AI uses and shares with a user cannot always carefully be controlled. Consider two information retrieval services connected to **two** databases. One is connected to a Structured Query Language (SQL) server. Another retrieves information semantically, using AI. With the retrieval service, the information being fetched and shared can be tightly controlled. With the semantic system, queries are based on semantics, and as such, it is more challenging to control what data is shared.<br><br>**Example Informative References:** DASF 5, 9, 11, 13, 15-16, 20-27, 29-34, 36, 42, 44-46, 56, 58-60, 62; ATLAS AML.M0024; OWASP (all); ENISA Threat Landscape 2025; https://arxiv.org/pdf/2303.00654; AI 100-2e2025 | **Sample Focus Area Considerations:** Minimize the use of sensitive data in AI prompts and features and implement the use of runtime redaction and guardrails. Prevent data leakage by implementing output filtering, pattern detection, and access control.<br><br>**Example Informative References DASF 25,34,40** ";  ENISA Threat Landscape 2025; https://arxiv.org/pdf/2303.00654; AI 100-2e2025 | **Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **PR.DS-11:** Backups of data are created, protected, maintained, and tested | **General Considerations**: No general considerations identified—see Focus Area Considerations. | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Maintain protected, regularly tested backups of critical AI assets, including validated model | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Maintain protected, regularly tested backups of critical AI assets, including validated model | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: CP-06; CP-09 | versions, clean training datasets, and configuration files, to ensure rapid recovery from data poisoning or model compromise.<br><br>**Example Informative References:** DASF 8, 15, 17, 20-22, 27-30, 41-42; ATLAS AML.M0012; ATLAS AML.M0024; ATLAS AML.M0025 | versions, clean training datasets, and configuration files, to ensure rapid recovery from data poisoning or model compromise.<br><br>**Example Informative References:** ENISA Threat Landscape 2025;DASF 17,38,50; ATLAS AML.M0014; ATLAS AML.M0007 | **Example Informative References:** NIST SP 800-53, Rev 5: CP-06, CP-09; |
| **Platform Security (PR.PS)** | The hardware, software (e.g., firmware, operating systems, applications), and services of physical and virtual platforms are managed consistent with the organization's risk strategy to protect their confidentiality, integrity, and availability | | | |
| **PR.PS-01:** Configuration management practices are established and applied | **General Considerations**: Configuration of hyperparameters to an AI system is critical to the functionality of the system. These configuration values are tracked, versioned, and managed with the same level of scrutiny afforded to the software itself.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-01; CM-02; CM-03; CM-04; CM-05; CM-06; CM-07; CM-08; CM-09; CM-10; CM-11 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI's performance is closely linked with its configuration settings. When adding AI to a broader system, these additional settings are also tracked and managed.<br><br>**Example Informative References:** DASF 3, 5-8, 10, 18-19, 21, 23-24, 28-29, 31-33, 35, 41-46, 51-53, 58, 60, 63, 64; ATLAS AML.M0005; ATLAS AML.M0012; OWASP Conventional Runtime Controls; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Establish and apply AI model configuration management practices, such as versioning and tracking model configurations, prompts, thresholds, and guardrail rules.<br><br>**Example Informative References:** DASF 38,40,41; ATLAS AML.M0023; ATLAS AML.M0021 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** NIST SP 800-218 (all) |
| **PR.PS-02:** Software is maintained, replaced, and removed commensurate with risk | **General Considerations**: Regularly check for and apply patches to AI frameworks and software libraries, both to defend the organization | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Regularly check for and apply patches to AI frameworks and libraries to close | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Un-secure, un-tested, and/or un-maintained code could be targeted by AI- |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | and to prevent attacks before they happen.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-11; MA-03(06); SA-10(01); SI-02; SI-07 | **Example Informative References:** DASF 13, 21, 23, 32, 38, 45, 53, 63; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | vulnerabilities that could be exploited during attacks.<br><br>**Example Informative References:** DASF 32,37,41; ENISA Threat Landscape 2025 | enabled cyber-attacks. Additional code scanning and replacement is established in accordance with vulnerability management to maintain risk thresholds. Follow procedures for maintaining software through its lifecycle.<br><br>**Example Informative References:** NIST SP 800-218 (all); NTIA SBOM (all); AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **PR.PS-03:** Hardware is maintained, replaced, and removed commensurate with risk | **General Considerations**: AI introduces a demand for new sets of hardware (accelerated compute). Ensure that this new hardware is regularly tested, maintained, and updated to minimize exposure to attacks.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-07 (09); SA-10(03); SC-03(01); SC-39(01); SC-49; SC-51 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 6, 10, 18-19, 29, 42, 59, 63; OWASP Conventional Runtime Controls; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Frequently update firmware and drivers for hardware accelerators.<br><br>**Example Informative References:** DASF 38-39; ENISA Threat Landscape 2025; ATLAS AML.M0016; AML.M0011 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Un-secure, un-tested, and/or un-maintained cyber physical hardware and systems could be targeted by AI-enabled attacks, additional hardware scanning and replacement is established to maintain risk thresholds.<br><br>**Example Informative References:** ATT&CK M1034 |
| **PR.PS-04:** Log records are generated and made available for continuous monitoring | **General Considerations**: AI introduces new distinct considerations across all focus areas.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-02; AU-03; AU-06; AU-07; AU-11; AU-12 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI presents a new set of metrics and parameters that need to be logged to have a complete picture of the system. For | **Proposed Priority:** 1<br><br>**Sample Opportunities:** Leverage AI to analyze and scan logs for anomalous events automatically. Utilize AI to compose information from multiple sources. Logs should show request data, response data, | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** The scale and speed are one such way AI-enabled attacks may challenge cybersecurity detection and monitoring systems—continuous |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | instance, detection thresholds and other inference time parameters.<br><br>**Example Informative References:** DASF 11-12, 14, 20-21, 23, 29, 32-33, 35-37, 39-40, 42, 44-46, 55, 60; ATLAS AML.M0005; OWASP Conventional Runtime Controls; OWASP GenAI Security Project: Monitor; ENISA Threat Landscape 2025; AI 100-2e2025 | and whether responses matched policy.<br><br>**Sample Focus Area Considerations:** Confirm findings with HITL reviews. Consider protecting logs from tampering.<br><br>**Example Informative References:** DASF 32,39,41; ENISA Threat Landscape 2025; AI 100-2e2025 | monitoring and logging will be needed to identify new patterns of AI-enabled tactics and techniques for detection and defense.<br><br>**Example Informative References:** AI 100-2e2025 |
| **PR.PS-05:** Installation and execution of unauthorized software are prevented | **General Considerations**: Do not allow automatic download or execution of code from any system. Ensure all software is scanned for malware and the source of the software is properly vetted.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CM-07(02); CM-07(04); CM-07(05); SC-34 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI agent systems are sometimes allowed to execute arbitrary code. In most applications, this ability should be curtailed, sandboxed, subject to approval and monitoring, or completely disallowed.<br><br>**Example Informative References:** DASF 7, 10-11, 13, 18-19, 29, 36, 44, 52, 54, 60; OWASP Conventional Runtime Controls; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Block unapproved model downloads or helpful agent scripts, ensuring that only vetted and authorized components are integrated into an AI defense environment.<br><br>**Example Informative References:** DASF 1, 40; ATLAS AML.M0011; ATLAS AML.M0013 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Un-authorized and un-scanned code is one such way that malicious AI-enabled packages and malware can be introduced into the system. Maintain strict guidance on downloading and installing software into production systems that may have access to other assets on the system.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **PR.PS-06:** Secure software development practices are integrated, and their performance is monitored throughout | **General Considerations**: AI introduces a new set of security metrics which extend beyond normal software development. These metrics are considered and | **Proposed Priority:** 2<br><br>**Focus Area Consideration**: AI presents a new set of performance considerations such | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Extend AI checks to secure software development | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| the software development life cycle | added to other security metrics the organization is tracking.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: SA-03; SA-08; SA-10; SA-11; SA-15; SA-15(13); SA-17; SA-24 | as repeatability, potential, bias towards certain inputs, and contingencies in the face of uncertainty model output.<br><br>**Example Informative References:** NIST SP 800-218A; DASF 11, 19-20, 23, 41, 45, 52, 59; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP Model Input Confidentiality; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2504.11168 ; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | practices by implementing prompt tests, adversarial tests, and evaluation gates; block deployment if safety or quality drops during performance monitoring. Apply insights learned from AI-enabled vulnerability scanning tools to secure AI systems and thwart AI-enabled attacks.<br><br>**Example Informative References:** DASF 38,41,45; ENISA Threat Landscape 2025; ATLAS AML.M0008; ATLAS AML.M0003; AI 100-2e2025 | **Example Informative References:** https://arxiv.org/pdf/2504.11168 |
| **Technology Infrastructure Resilience (PR.IR)** | Security architectures are managed with the organization's risk strategy to protect asset confidentiality, integrity, and availability, and organizational resilience | | | |
| **PR.IR-01:** Networks and environments are protected from unauthorized logical access and usage | **General Considerations**: AI's ability to access new networks or systems within a network should be curtailed commensurate with the functionality of the AI system and the level of risk the organization is willing to take.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-03; AC-04; SC-04; SC-05; SC-07 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI as an agent can attempt to access networks and environments distinctly from previously seen agents (people and classically controlled bots). AI agents' access to networks and environments should be restricted by the principle of least | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Gate AI-initiated network changes and privilege use by implementing human approval or policy-based checks.<br><br>**Example Informative References:** DASF 38,40; ENISA Threat Landscape 2025; ATLAS AML.M0019; AI 100-2e2025; | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI-enabled attacks will lower the barrier of entry to gaining access to networks and unauthorized access and usage. Defenses should prioritize additional layers of security for endpoint detection systems and credential and |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | privilege and according to the organization's risk strategy.<br><br>**Example Informative References:** DASF 1-5, 7, 9-11, 16, 18-19, 21, 23-24, 28-34, 41, 43, 45-46, 52, 56-59, 64; ATLAS AML.M0005; ATLAS AML.M0012; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP AI Exchange: Rate Limit; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2504.11168 ; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | https://arxiv.org/html/2503.11917v3 | identity management such as MFA and Zero-Trust architectures.<br><br>**Example Informative References:** AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |
| **PR.IR-02:** The organization's technology assets are protected from environmental threats | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-02; PE-09; PE-10; PE-11; PE-12; PE-13; PE-14; PE-15; PE-18; PE-23 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 23, 45 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 17, 41; ATLAS AML.M0009; ATLAS AML.M0005 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI-enabled attacks can impact physical assets (e.g., operational technologies).<br><br>**Example Informative References:** AI 100-2e2025 |
| **PR.IR-03:** Mechanisms are implemented to achieve resilience | **General Considerations**: Various systems exist to increase the resilience of AI models including but not limited to backups, model | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** It may be | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Implement AI- | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** The speed and |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| requirements in normal and adverse situations | hardening, ensemble methods, and automated failover.<br><br>Each system has its own limitations which are understood. These limitations inform proper implementation.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP; IR; SA-08; SA-24; SC-06; SC-24; SC-36; SC-39; SI-13; | difficult to create backup systems or recreate data in the event of AI failure. If an issue is identified with a model, and that issue exists in previous models as well (such as being susceptible to adverse input), it may be non-trivial to create and deploy a new model resilient to the newly discovered issue.<br><br>**Example Informative References:** DASF 3, 5, 19, 24, 34; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP Model Input Confidentiality; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | specific resilience mechanisms, such as model hardening (e.g., adversarial training), ensemble methods, and automated failover to trusted model versions, to ensure continuous defense operations during an adverse event.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 38,39; ATLAS AML.M0003; ATLAS AML.M0006; AI 100-2e2025 | scale of AI-enabled cyber attacks increase the importance of implementing resiliency controls.<br><br>**Example Informative References:** AI 100-2e2025 |
| **PR.IR-04:** Adequate resource capacity to ensure availability is maintained | **General Considerations**: AI, specifically deep neural networks, can be some of the most intensive computation workloads that are performed on a day-to-day basis. Several strategies can be used to get more out of limited resources (load balancing, queues, parallelism, etc.) but for critical applications, all these strategies have drawbacks and instead it is appropriate to have dedicated | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Organizations account for AI resource needs as workloads can be several orders of magnitude more resource-intensive than other non-AI compute workloads.<br><br>**Example Informative References:** DASF 5, 21, 23-24, v30, 34, 45; OWASP AI Exchange: General | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI forecasts hardware failures and system degradation that may affect defensive readiness.<br><br>**Sample Focus Area Considerations:** Due to the resource-intensive nature of AI workloads, reserve compute for AI-defense actions during incidents. | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI-enabled attacks may increase the resources required to properly detect when and if those attacks have occurred or are occurring. This is especially important when AI is implemented in cyber defense capabilities. See Defend for more considerations. |

| CSF 2.0 Core: PROTECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | compute infrastructure to applications which require high availability and low latency responses.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-06; CP-07; CP-08; PM-03; PM-09 | Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | **Example Informative References:** DASF 34,39,50; ENISA Threat Landscape 2025; ATLAS AML.M0004; ATLAS AML.M0017; AI 100-2e2025 | **Example Informative References:** AI 100-2e2025 |

## 2.6. Cyber AI Profile: DETECT

**Table 4 Cyber AI Profile – DETECT.**

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **DETECT (DE)** | Possible cybersecurity attacks and compromises are found and analyzed | | | |
| **Continuous Monitoring (DE.CM)** | Assets are monitored to find anomalies, indicators of compromise, and other potentially adverse events | | | |
| **DE.CM-01:** Networks and network services are monitored to find potentially adverse events | **General Considerations**: AI introduces new distinct considerations when securing applications and using AI to defend organizations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-02; AU-12; CA-07; CM-03; SC-05; SC-07; SI-04 | **Proposed Priority: 2**<br><br>**Sample Focus Area Considerations:** AI system traffic should be tracked, logged, and monitored separately from other network traffic. If an adverse event occurs, it is easy to identify the source as either human or AI initiated.<br><br>**Example Informative References:** DASF 3-5, 7, 9-11, 13, 18-19, 23-24, 29-31, 36, 41, 44-46, 52, 55-56, 60, 62; ATLAS AML.M0012; ATLAS AML.M0024; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP AI Exchange: Rate Limit; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025; | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting unusual patterns faster than humans and other automated tools.<br><br>AI agents, either separately or as a team, can monitor networks, validate defensive actions, and improve detection accuracy while reducing human workload.<br><br>Leverage AI to automatically track and monitor network traffic (e.g., conducting nmap scans) in support of detecting potentially adverse events.<br><br>Capitalize on new monitoring and analysis capabilities (e.g., user and machine behavioral analysis, real-time monitoring and response) and identify AI capabilities that can support organizations in | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend for additional guidance.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | https://arxiv.org/html/2503.11917v3 | moving from a reactive defensive mode to proactive (e.g., predictive analysis).<br><br>**Sample Focus Area Considerations:** Monitor and manage agents for continued effectiveness over time.<br><br>**Example Informative References: DASF 25,32,39,61;** ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | |
| **DE.CM-02:** The physical environment is monitored to find potentially adverse events | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:**<br><br>NIST SP 800-53, Rev 5: CA-07; PE-03; PE-06; PE-20 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI systems often sit in GPU clusters or shared racks. Monitor power, heat, and physical access, as tampering with these attributes can disable defenses.<br><br>**Example Informative References: DASF 32,39,41;** ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **DE.CM-03:** Personnel activity and technology usage are monitored to find potentially adverse events | **General Considerations**: No general considerations identified – see Focus Area Considerations.<br><br>**Example Informative References:**<br><br>NIST SP 800-53, Rev 5: AC-02; AU-12; AU-13; CA-07;<br><br>CM-10; CM-11 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Personnel usage of internal and external AI tools should be monitored to find adverse events or sharing of sensitive data.<br><br>**Example Informative References:** DASF 5, 7, 10-11, 13, 18-19, 23- | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI helps analysts to solve problems by handling common queries and generating alert summaries, and routing issues to the correct analyst.<br><br>AI agents, either separately or as a team, can monitor network, | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Personnel may be subject to AI-enabled phishing or deepfake attacks that could introduce malware onto the system. |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | 24, 29, 33, 36, 41-42, 44-46, 52-53, 55, 64; ATLAS AML.M0024; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025 | validate defensive actions, and improve detection accuracy while reducing human workload.<br><br>AI catches credential misuse that rules might miss by flagging unusual authentication activity.<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References: DASF 25,37,39;** ENISA Threat Landscape 202; AI 100-2e2025 | **Example Informative References:** https://arxiv.org/pdf/2305.06972; AI 100-2e2025 |
| **DE.CM-06:** External service provider activities and services are monitored to find potentially adverse events | **General Considerations**: Some AI applications may rely on third party services. Use logging to show requests, responses, metadata, and relevant metrics.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CA-07; PS-07; SA-04; SA-09; SI-04 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Many organizations are relying on external service providers for AI services. Due to this, the importance of monitoring external service providers remains high. Because the methods of monitoring these service providers remain the same, there are no AI-specific considerations.<br><br>**Example Informative References:** DASF 7, 10-11, 13, 18-19, 29, 36, 41, 44, 52, 60; ATLAS AML.M0024; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Many defense tools call third-party AI APIs. These APIs should be monitored to identify potentially adverse events.<br><br>**Example Informative References: DASF 25,39,41;** ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Third-party services and providers may introduce new vulnerabilities with updates and/or patches to software and systems that AI-enabled cyber attacks could identify and exploit. Additionally, as the number of third-party providers grows, the larger the threat landscape for AI-enabled cyber attacks to leverage without proper management and monitoring services and activities.<br><br>**Example Informative References:** AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | GenAI Security Project: Monitor; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | | |
| **DE.CM-09:** Computing hardware and software, runtime environments, and their data are monitored to find potentially adverse events | **General Considerations**: Because AI can operate autonomously, ensure that all internal AI systems are properly monitored for anomalous activity.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AC-04; AC-09; AU-12; CA-07; CM-03; CM-06; CM-10; CM-11; SC-34; SC-35; SI-04; SI-07 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Because AI can autonomously create and augment data as well as create and execute its own code, new monitoring is needed to track actions taken by AI. Inputs to and outputs from AI systems may be monitored to detect adverse events, such as monitoring for adversarial inputs or anomalous AI system behaviors. For human initiated AI activities, standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 5, 7-8, 10-13, 18-20, 23-24, 29-31, 36, 4-42, 44-46, 51-53, 55-56, 60; ATLAS AML.M0012; ATLAS AML.M0024; OWASP Controls to limit the effects of the unwanted behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; OWASP AI Exchange: Model Access Control; OWASP GenAI Security Project: Monitor; | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting unusual patterns faster than humans and other automated tools.<br><br>A team of AI agents monitors network, validate defensive actions, and improve detection accuracy while reducing human workload.<br><br>**Sample Focus Area Considerations:** Monitor AI systems and their runtime environments for anomalous behaviors (e.g., unexpected file writes, API calls, or generated binaries) that may indicate adversarial manipulation, data exfiltration, or model exploitation.<br><br>**Example Informative References: DASF 25,39,41;** ENISA Threat Landscape 2025; ATLAS | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI-enabled cyber attacks may leverage local compute resources to accomplish their objectives. Monitor resources and runtime environments for anomalous and/or adverse events.<br><br>**Example Informative References:** AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | AML.M0024; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | |
| **Adverse Event Analysis (DE.AE)** | Anomalies, indicators of compromise, and other potentially adverse events are analyzed to characterize the events and detect cybersecurity incidents | | | |
| **DE.AE-02:** Potentially adverse events are analyzed to better understand associated activities | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-06; CA-07; IR-04; SI-04 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Potentially adverse events impacting AI systems, such as instances of adversarial inputs, are analyzed to characterize the event and associated activities.<br><br>**Example Informative References:** DASF 7, 12-14, 16, 19, 23, 29, 31, 35-339, 42, 44, 46, 55-56, 59; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting unusual patterns faster than human and other automated tools.<br><br>AI helps analysts to solve harder problems by handling common queries and generating alert summaries, and routing issues to the correct analyst).<br><br>**Sample Focus Area Considerations:** When analyzing potentially adverse events, human review is needed to supplement AI-defense actions and to avoid noise chasing.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 25,32,39; | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Analyzing potentially adverse events for indicators of AI usage can aid organizations in understanding potential AI-enabled threats. Standard cybersecurity practices apply. See Defend for additional guidance.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | https://arxiv.org/html/2503.11917v3 | |
| **DE.AE-03:** Information is correlated from multiple sources | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-06; CA-07; PM-16; IR-04; IR-05; IR-08; SI-04 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12-14, 16, 19, 23, 29, 31, 35-37, 39, 42, 44, 46, 55-56, 59-60; ATLAS AML.M0024; OWASP Conventional Runtime Controls; OWASP AI Exchange: General Governance Controls; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting unusual patterns faster than humans and other automated tools.<br><br>A team of AI agents monitors network, validate defensive actions, and improve detection accuracy while reducing human workload.<br><br>**Sample Focus Area Considerations:** The aggregation of data from multiple log sources enhances AI-cyber defenses in detecting anomalous and potentially adverse events.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 32,39,41; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity patches apply.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |
| **DE.AE-04:** The estimated impact and scope of adverse events are understood | **General Considerations**: No general considerations identified—see Focus Area Considerations. | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Scope estimates should be cross-checked with | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Understanding the potential scope and scale of |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: PM-09; PM-11; PM-18; PM-28; PM-30 | **Example Informative References:** DASF 13, 16, 19, 23, 31, 39, 56, 59-60; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain | ground truth data with human review AI generated for accuracy.<br><br>**Example Informative References:** DASF 32,39,41 | AI-enabled cyber attacks better positions the organization to implement prevention and resiliency measures.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **DE.AE-06:** Information on adverse events is provided to authorized staff and tools | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; PM-15; PM-16; RA-04; RA-10 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12-14, 16, 19, 23, 29, 35-39, 42, 44, 46, 55-56, 59; ATLAS AML.M0024; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Explain and share AI findings related to adverse events, including context and confidence scores.<br><br>**Example Informative References:** DASF 25,32,39; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI-enabled cyber attacks may impact multiple security and software teams, in addition to third-party providers and tools. To aid in prioritization and scoped remediation, all teams should be aware of adverse events.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **DE.AE-07:** Cyber threat intelligence and other contextual information are integrated into the analysis | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: PM-16; RA-03; RA-10 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12-14, 16, 19, 23, 29, 31, 35-37, 39, 42, 44, 46, 55-56, 60-61; OWASP AI Exchange: General Governance Controls; OWASP | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI improves monitoring and strengthens threat detection by flagging anomalies, correlating suspicious behaviors, and spotting unusual patterns faster than humans and automated tools. | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend for additional guidance.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | LLM Top Ten: LLM03 Supply Chain; https://arxiv.org/html/2503.11917v3 | AI helps defenders by scanning massive datasets and rapidly analyzing large volumes of threat intelligence to give defenders a clear view of the threat landscape.<br><br>AI collects, formats, and shares threat intelligence using standards protocols (such as STIX and OpenCTI) to ensure teams and partners stay aligned.<br><br>**Sample Focus Area Considerations:** Feed adversarial ML research and AI red-team reports into detection pipelines to enhance detection capabilities and support response and recovery efforts.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 25; https://arxiv.org/html/2503.11917v3 | |
| **DE.AE-08:** Incidents are declared when adverse events meet the defined incident criteria | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 16, 19, 23, 39, 56, 59; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References: DASF 25,32,39;** ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Adjust thresholds and maintain explainable escalation criteria for declaring an AI-enabled cyber attack.<br><br>**Example Informative References:** AI-specific Example Informative |

| CSF 2.0 Core: DETECT | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | | References pending additional inputs. |

## 2.7. Cyber AI Profile: RESPOND

**Table 5 Cyber AI Profile - RESPOND.**

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RESPOND (RS)** | Actions regarding a detected cybersecurity incident are taken | | | |
| **Incident Management (RS.MA)** | Responses to detected cybersecurity incidents are managed | | | |
| **RS.MA-01:** The incident response plan is executed in coordination with relevant third parties once an incident is declared | **General Considerations**: Communicate with third party service and data providers.<br><br>**Example Informative References:**<br><br>NIST SP 800-53, Rev 5: IR-06; IR-07; IR-08; SR-03; SR-08 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** In instances where third parties are being used for AI services, it may be appropriate to coordinate with those parties when responding to an incident to ensure the appropriate incident response activities are conducted and appropriately coordinated between parties. Incidents related to data leakage may have additional considerations based on the nature of the data (e.g., leaking information that can impact an individual's privacy may trigger breach notification | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI supports automation for incident responses like triage and recommending containment steps so teams can move faster and with greater confidence during an incident.<br><br>**Sample Focus Area Considerations:** If defensive AI actions rely on outside APIs or datasets, coordinate with those providers during incidents.<br><br>**Example Informative References:** DASF 25,39,41; ENISA Threat Landscape 2025; AI 100-2e2025; | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** The speeds and scale with which AI-enabled attacks occur require more rapid response and coordination with any relevant third parties.<br><br>**Example Informative References:** AI 100-2e2025; https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | **Secure** | **Defend** | **Thwart** |
| | | requirements under various privacy laws and regulations).<br><br>**Example Informative References:** DASF 39; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025; https://arxiv.org/html/2503.11917v3 | https://arxiv.org/html/2503.11917v3 | |
| **RS.MA-02:** Incident reports are triaged and validated | **General Considerations**: AI-related reports and AI-related activities should be categorized differently from other activities/reports.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-05; IR-06 | **Proposed Priority**: 2<br><br>**Sample Focus Area Considerations:** AI presents new threat vectors and attack methods. Establish criteria for triaging and validating AI-related incidents.<br><br>**Example Informative References:** DASF 7, 12-14, 16, 29, 35-38, 42, 44, 46, 55-56, 60; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI supports automation for incident responses like triage and recommends containment steps so teams can move faster and with greater confidence during an incident.<br><br>**Sample Focus Area Considerations:** Incident reports should be updated to reflect the potential use of AI-enabled cyber defense actions during incident response. Human review and validation are included in the process to ensure appropriate actions are taken.<br><br>**Example Informative References:** DASF 12,37,39; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See the Defend Focus Area Considerations for additional guidance when AI-enabled defenses are integrated into incident response pipelines such as for collecting incident data, compiling, validating, and providing reports.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RS.MA-03:** Incidents are categorized and prioritized | **General Considerations**: AI introduces new distinct considerations when securing applications and when using AI to defend organizations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-05 | **Proposed Priority**: 2<br><br>**Sample Focus Area Considerations:** Because AI presents new threat vectors and attack methods, new categorizations need to be created for incidents which capture this added dimension.<br><br>**Example Informative References:** DASF 13, 39; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Integrate AI-driven analytics into incident categorization and prioritization to identify and flag AI-influenced events (e.g., adversarial attacks, data poisoning) in real time.<br><br>**Example Informative References:** DASF 25,39,41; ENISA Threat Landscape 2025; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See the Defend Focus Area Considerations for additional guidance.<br><br>**Example Informative References:** ATLAS AML.M0015; https://arxiv.org/html/2503.11917v3 |
| **RS.MA-04:** Incidents are escalated or elevated as needed | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-05; IR-06; IR-07 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 13, 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI can facilitate decision-making regarding when to escalate or elevate incidents.<br><br>**Sample Focus Area Considerations:** Review results of AI incident classifications. If AI misclassifies an incident or incorrectly amplifies its impacts, flag it early and escalate to human analysts.<br><br>**Example Informative References:** DASF 25,32,39; ENISA Threat Landscape 2025; ATLAS AML.M0017 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Implement automated actions such as blocking and/or isolating systems in addition to AI-enabled defenses to flag adverse behaviors and events for additional review.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RS.MA-05:** The criteria for initiating incident recovery are applied | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Recovery may include retraining or disabling AI modules. Define triggers for when to roll back an AI component.<br><br>**Example Informative References:** DASF 38,39; ENISA Threat Landscape 2025; ATLAS AML.M0008; ATLAS AML.M0014; https://arxiv.org/html/2503.11917v3 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend for additional guidance such as implementing AI to help identify and restore systems.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **Incident Analysis (RS.AN)** | Investigations are conducted to ensure effective response and support forensics and recovery activities | | | |
| **RS.AN-03:** Analysis is performed to establish what has taken place during an incident and the root cause of the incident | **General Considerations**: New analysis tools may be needed to capture the entire picture of AI system attacks and AI system defenses.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-07; IR-04; SI-02(07) | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** New expertise, tools, and methods may be needed to diagnosis more complex attacks on AI such as adversarial input.<br><br>**Example Informative References:** DASF 39, 55; OWASP Conventional Runtime Controls; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Analyze AI-specific artifacts (e.g., model logs, inference tables, provenance data) during an incident to establish the root cause.<br><br>**Example Informative References:** DASF 21, 37; ATLAS AML.M0024 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Incident analysis explicitly searches for indicators of adversary AI usage in the incident. AI-enabled attacks may have unique indicators or signatures due to their speed and scale, and their dynamic and optimized nature. Adversary use of AI is rapidly evolving and organizations may need to track cyber threat intelligence closely in order to determine if/how AI was used in the incident.<br><br>**Example Informative References:** AI-specific Example Informative |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | | References pending additional inputs. |
| **RS.AN-06:** Actions performed during an investigation are recorded, and the records' integrity and provenance are preserved | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-07; IR-04; IR-06 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39, 55; OWASP Conventional Runtime Controls; OWASP GenAI Security Project: Monitor; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 25,39,41, ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** Incident response activities and findings inform future efforts to build resiliency and support overall improvement efforts (see also ID.IM-04).<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **RS.AN-07:** Incident data and metadata are collected, and their integrity and provenance are preserved | **General Considerations**: AI may result in capturing new types of metadata (e.g., data versions, inputs, hyperparameters).<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: AU-07; IR-04; IR-06 | **Proposed Priority:** 1<br><br>**Sample Focus Area Considerations:** AI introduced the need for dataset tracking and versioning as well as documentation of associated metadata related to the model, such as hyperparameters. The parameters used to train the model, when it was trained, and what version of the dataset it used are also relevant.<br><br>**Example Informative References:** OWASP Conventional Runtime Controls; OWASP GenAI Security Project: Monitor; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Preserve logs, inputs, outputs, and decision chains of AI systems to ensure provenance and improving future AI-driven response actions.<br><br>**Example Informative References: DASF 25,39,41;** ENISA Threat Landscape 2025; ATLAS AML.M0024 (AI Telemetry Logging) | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RS.AN-08:** An incident's magnitude is estimated and validated | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-08; RA-03; RA-07 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 13, 39; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Estimate and validate the magnitude of an AI-influenced incident by assessing the scope of adversarial impact on model integrity, the quantity of exposed sensitive data (e.g., training data leakage), and the duration of model availability loss.<br><br>**Example Informative References:** DASF 25,32,41; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. Understanding the magnitude of actual incidents informs priorities for implementing future resiliency measures. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **Incident Response Reporting and Communication (RS.CO)** | Response activities are coordinated with internal and external stakeholders as required by laws, regulations, or policies | | | |
| **RS.CO-02:** Internal and external stakeholders are notified of incidents | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-06; IR-07; SR-03; SR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12, 14, 29, 35-37, 39, 42, 46, 55; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Notification processes are established for AI-related incidents (e.g., adversarial attacks, unexpected model failures) to rapidly inform internal teams and external suppliers of incidents, such as compromised models or data. Establish criteria for determining when incident notifications can be autonomously delivered and when they require human review first (e.g., notifications regarding incidents | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | that may have legal or compliance implications may require human review before they are sent).<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 25, 41; ENISA Threat Landscape 2025 ATLAS AML.M0024; | |
| **RS.CO-03:** Information is shared with designated internal and external stakeholders | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-06; IR-07; SR-03; SR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area**<br><br>**Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 7, 12, 14, 29, 35-37, 39, 42, 46, 55; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI supports reporting by drafting clear incident summaries and compliance reports, organizing evidence, and producing standardized documentation to recuse workload of analysts.<br><br>AI formats, and shares threat intelligence using standards protocols (such as STIX and OpenCTI) to ensure teams and partners stay aligned.<br><br>AI translates complex technical data into clear business language by summarizing technical risks into concise insights and helping executives make informed risk decisions.<br><br>**Sample Focus Area Considerations:** Establish protocols for sharing AI-specific threat intelligence and incident metrics (e.g., adversarial inputs) | **Proposed Priority:** 3<br><br>**Sample Focus Area**<br><br>**Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | with internal developers and external partners for coordinated defense hardening.<br><br>**Example Informative References: DASF 25,32,41;** ENISA Threat Landscape 2025 | |
| **Incident Mitigation (RS.MI)** | Activities are performed to prevent expansion of an event and mitigate its effects | | | |
| **RS.MI-01:** Incidents are contained | **General Considerations**: No general considerations identified - see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. (Rationale) Because AI-initiated incidents can move at faster-than-human speeds, the importance of containing such incidents is elevated. However, the methods of containing these incidents remain the same.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025; https://arxiv.org/pdf/2303.00654 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI supports automation for incident responses like triage and recommending containment steps, so teams can move faster during an incident.<br><br>**Sample Focus Area Considerations:** Containment procedures include AI-specific steps, such as rapidly disabling the autonomy/privileges of a compromised AI agent to a validated, trusted state.<br><br>**Example Informative References: DASF 25,32,39,41;** ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** AI-enabled cyber-attacks could adapt to avoid detection or create patterns of attacks that make detection difficult—leverage mitigations to block and/or isolate systems to ensure AI-enabled attacks cannot proliferate within the system. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **RS.MI-02:** Incidents are eradicated | **General Considerations**: No general considerations identified—see Focus Area Considerations. | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI supports automation for incident responses, like triage and recommending containment steps, | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. |

| CSF 2.0 Core: RESPOND | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | **Example Informative References:** NIST SP 800-53, Rev 5: IR-04 | **Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | so teams can move faster during an incident.<br><br>**Sample Focus Area Considerations:** Eradication addresses the root cause of AI-enabled compromise, including poisoned training data, patching vulnerable AI libraries, and fully revoking access from adversarial tools.<br><br>**Example Informative References:** DASF 25,32,41; ENISA Threat Landscape 2025 | **Example Informative References:** AI-specific Example Informative References pending additional inputs. |

## 2.8. Cyber AI Profile: RECOVER

**Table 6 Cyber AI Profile - RECOVER.**

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RECOVER (RC)** | Assets and operations affected by a cybersecurity incident are restored | | | |
| **Incident Recovery Plan Execution (RC.RP)** | Restoration activities are performed to ensure operational availability of systems and services affected by cybersecurity incidents | | | |
| **RC.RP-01:** The recovery portion of the incident response plan is executed once initiated from the incident response process | **General Considerations**: Recovery of an AI-related incident may not be straightforward. Extra considerations for AI may be needed. The complexity of recovery depends on the type of | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Recovery efforts for AI systems introduces complications that may not be necessary for recovery efforts for | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI accelerates recovery by calculating which systems to restore first, tracking progress, and drafts clear | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Recovery may require software patches and/or updates, change of passwords, installation of new or updated |

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | attack, the type of model, and the severity of the incident instance. In the simplest of cases, recovery may consist of a data restoration from backups, a software update, or updating credentials. In more complicated cases, an organization may need to roll back to older versions of the software system. In the most complicated of cases, it may require completely retraining the model.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-10; IR-04; IR-08 | other types of systems (e.g., retraining a model).<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | updates to keep stakeholders informed.<br><br>**Sample Focus Area Considerations:** Recovery activities include evaluating whether AI-related cyber defense capabilities have returned to a reliable state. Recovery may require retraining models or rolling back to a safer checkpoint.<br><br>**Example Informative References:** DASF 38,39,50; ENISA Threat Landscape 2025; ATLAS AML.M0008; https://arxiv.org/pdf/2505.14835 | anti-virus and/or firewall protections.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **RC.RP-02:** Recovery actions are selected, scoped, prioritized, and performed | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-10; IR-04; IR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 1<br><br>**Sample Opportunities:** AI accelerates recovery by calculating which systems to restore first, tracking progress, and drafts clear updates to keep stakeholders informed.<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 38,39,50; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RC.RP-03:** The integrity of backups and other restoration assets is verified before using them for restoration | **General Considerations**: Regularly conduct tests on backups and previous models to detect concept drift or other forms of model degradation.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-02; CP-04; CP-09 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Model data may be responsible for an incident. In this case, data may need to be changed or augmented to patch the vulnerability.<br><br>**Example Informative References:** DASF 8, 31, 41; ATLAS AML.M0012; AI 100-2e2025 | **Proposed Priority:** 2<br>**Sample Focus Area Considerations:** Assure the integrity of AI components by testing model and dataset backups for poisoning or drift.<br><br>**Example Informative References:** DASF 43,47; AI 100-2e2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **RC.RP-04:** Critical mission functions and cybersecurity risk management are considered to establish post-incident operational norms | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example**<br><br>**Informative References:** NIST SP 800-53, Rev 5: PM-08; PM-09; PM-11; IR-01; IR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 2<br><br>**Sample Opportunities:** AI forecasts hardware failures and system degradation that may affect defensive readiness.<br><br>**Sample Focus Area Considerations:** After an incident, evaluate how AI defense systems performed (e.g., detecting accuracy, FPs) and determine how AI can assist in refining post-incident operational norms<br><br>**Example Informative References:** DASF 39,50; ENISA Threat Landscape 2025; OWASP AI Exchange: How about Responsible or Trustworthy AI?; ATLAS AML.M0008; | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| **RC.RP-05:** The integrity of restored assets is verified, systems and services are restored, and normal operating status is confirmed | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-10 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 31 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI accelerates recovery by calculating which systems to restore first, tracking progress, and drafts clear updates to keep stakeholders informed.<br><br>**Sample Focus Area Considerations:** Verify the integrity of restored AI components (models, training data) for compromise (e.g., residual poisoning) and validate that the restored AI defense system operates at expected performance (e.g., model accuracy, FP rate) before confirming normal operational status.<br><br>**Example Informative References:** DASF 43,47; ATLAS AML.M0014; ATLAS AML.M0008; ENISA Theat Landscape 2025; https://arxiv.org/pdf/2505.14534v1 | **Proposed Priority:** 2<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend Considerations for additional guidance.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **RC.RP-06:** The end of incident recovery is declared based on criteria, and incident-related documentation is completed | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** ENISA Theat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** Conduct reporting (e.g., writing after action reports and client monitoring outputs).<br><br>**Sample Focus Area Considerations:** Documentation | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply. See Defend Considerations for additional guidance. |

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | includes AI-specific artifacts (e.g., model logs, provenance records) and a final post-mortem detailing how AI defense systems performed and how adversarial vectors (e.g., prompt injection) were mitigated or defeated.<br><br>**Example Informative References:** DASF 39,50; ATLAS AML.M0024; ATLAS AML.M0025; ENISA Theat Landscape 2025 | **Example Informative References:** AI-specific Example Informative References pending additional inputs. |
| **Incident Recovery Communication (RC.CO)** | Restoration activities are coordinated with internal and external parties | | | |
| **RC.CO-03:** Recovery activities and progress in restoring operational capabilities are communicated to designated internal and external stakeholders | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: IR-04; IR-06; SR-08 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 31, 39; OWASP AI Exchange: General Governance Controls; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Opportunities:** AI accelerates recovery by calculating which systems to restore first, tracking progress, and drafts clear updates to keep stakeholders informed.<br><br>**Sample Focus Area Considerations:** Communication during AI-related recovery informs stakeholders of the model's status (e.g., restored model version) and the expected delay in reenabling automated defense capabilities.<br><br>**Example Informative References:** ENISA Threat Landscape 2025; DASF 39; | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** https://arxiv.org/html/2503.11917v3 |

| CSF 2.0 Core: RECOVER | General Considerations | Focus Area Proposed Priorities & Considerations | | |
|---|---|---|---|---|
| | | Secure | Defend | Thwart |
| | | | https://arxiv.org/html/2503.11917v3 | |
| **RC.CO-04:** Public updates on incident recovery are shared using approved methods and messaging | **General Considerations**: No general considerations identified—see Focus Area Considerations.<br><br>**Example Informative References:** NIST SP 800-53, Rev 5: CP-02; IR-04; | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 31, 39; ENISA Threat Landscape 2025 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** DASF 25,32,39,41; ENISA Threat Landscape 2025; ATLAS AML.M0002 | **Proposed Priority:** 3<br><br>**Sample Focus Area Considerations:** Standard cybersecurity practices apply.<br><br>**Example Informative References:** AI-specific Example Informative References pending additional inputs. |

## References

The following references informed development of the Cyber AI Profile:

[1] Barbosa, D. O., Burbano, L., Yang, S., Wang, Z., Cardenas, A. A., Xie, C., & Cao, Y. (2025). *Robust and Efficient AI-Based Attack Recovery in Autonomous Drones.* arXiv. https://doi.org/10.48550/arXiv.2505.14835

[2] Bernardez Molina, S., Gómez Mármol, F., & Nespoli, P. (2024). *Tackling Cyberattacks through AI-based Reactive Systems: A Holistic Review and Future Vision.* arXiv. https://doi.org/10.48550/arXiv.2312.06229

[3] Apruzzese, G., Anderson, H. S., Dambra, S., Freeman, D., Pierazzi, F., & Roundy, K. (2023). *"Real Attackers Don't Compute Gradients": Bridging the Gap Between Adversarial ML Research and Practice.* Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P 2023). University of Liechtenstein; Robust Intelligence; Norton Research Group; Meta; King's College London. arXiv. https://doi.org/10.48550/arXiv.2212.14315

[4] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Cevik, M., & others. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* Future of Humanity Institute, University of Oxford; Centre for the Study of Existential Risk, University of Cambridge; Center for a New American Security; Electronic Frontier Foundation; and OpenAI. arXiv. https://doi.org/10.48550/arXiv.1802.07228

[5] Kiribuchi, N., Zenitani, K., & Semitsu, T. (2025). *Securing AI Systems: A Guide to Known Attacks and Impacts.* Japan AI Safety Institute (J-AISI) and Information-technology Promotion Agency (IPA), Tokyo, Japan. arXiv. https://doi.org/10.48550/arXiv.2506.23296

[6] Databricks (2025). *Databricks AI Security Framework (DASF) Version 2.0.* https://www.databricks.com/resources/whitepaper/databricks-ai-security-framework-dasf

[7] The MITRE Corporation (2025). MITRE ATLAS™. https://atlas.mitre.org/

[8] National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0) (National Institute of Standards and Technology, Gaithersburg, MD), NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

[9] European Union Agency for Cybersecurity (2023). Multilayer Framework for Good Cybersecurity Practices for AI. https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for-ai

[10] Open Worldwide Application Security Project (2024). OWASP Top 10 for LLM Applications 2025. https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/

[11] M. Rodriguez. et. al. (2025). *A Framework for Evaluating Emerging Cyberattack Capabilities of AI*. arXiv. https://arxiv.org/html/2503.11917v3#S5

[12] M. Bhtt . et. al. (2023). *Purple Llama CyberSecEval: A Secure Coding Benchmark for Language Models.* arXiv. https://doi.org/10.48550/arXiv.2312.04724

[13] M. Kouremetis. et. Al. (2025). *OCCULT: Evaluating Large Language Models for Offensive Cyber Operation Capabilities.* arXiv. https://doi.org/10.48550/arXiv.2502.15797.
[14] Anthropic (2025). *Disrupting the first reported AI-orchestrated cyber espionage campaign.* https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf
[15] Open Worldwide Application Security Project (2025). OWASP AI Exchange. https://owaspai.org/OWASP-AI-Exchange.pdf
[16] Open Worldwide Application Security Project (2025). OWASP AI Testing Guide. https://owasp.org/www-project-ai-testing-guide/
[17] National Institute of Standards and Technology (2025). NIST AI 100-2 Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. https://csrc.nist.gov/pubs/ai/100/2/e2025/final
[18] Bumanglang, K; Flossman, M; Hartley, N; Nimmo, B.; Stubbs, J; & Zhang, A (2025). *Disrupting malicious uses of AI: October 2025*. OpenAI. https://cdn.openai.com/threat-intelligence-reports/7d662b68-952f-4dfd-a2f2-fe55b041cc4a/disrupting-malicious-uses-of-ai-october-2025.pdf
[19] Walker, K (2025). *A summer of security: empowering cyber defenders with AI.* Google. https://blog.google/technology/safety-security/cybersecurity-updates-summer-2025/
[20] Waisman, N (2025). *XBOW on HackerOne: What's Next.* XBOW. https://xbow.com/blog/xbow-on-hackerone-whats-next
[21] Ullah, S. et. al. (2025). *From CVE Entries to Verifiable Exploits: An Automated Multi-Agent Framework for Reproducing CVEs.* arXiv. https://arxiv.org/abs/2509.01835v1
[22] National Institute of Standards and Technology (2024). The NIST Cybersecurity Framework (CSF) 2.0 (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) 29. https://doi.org/10.6028/NIST.CSWP.29
[23] Joint Task Force Transformation Initiative (2020). Security and Privacy Controls for Information Systems and Organizations (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-53, Rev. 5. https://doi.org/10.6028/NIST.SP.800-53r5
[24] Zhao, H. et al. (2023). *Explainability for Large Language Models: A Survey.* arXiv. https://arxiv.org/abs/2309.01029
[25] European Union Agency for Cybersecurity (2025). ENISA Threat Landscape 2025. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025
[26] Boyens, J; Smith A; Bartol N; Winkler K; Holbrook, A; & Fallon M. (2022). Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-161r1-upd1. https://doi.org/10.6028/NIST.SP.800-161r1-upd1
[27] Ross R; Pillitteri, V; Guissanie, G; Wagner, R; Graubart, R; & Bodeau, D (2021). Enhanced Security Requirements for Protecting Controlled Unclassified Information (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-72. https://doi.org/10.6028/NIST.SP.800-172

[28] United States Department of Commerce (2021). The Minimum Elements for a Software Bill of Materials (SBOM). https://www.ntia.gov/report/2021/minimum-elements-software-bill-materials-sbom

[29] Souppaya, M; Scarfone, K; & Dodson, D (2022). Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-218. https://doi.org/10.6028/NIST.SP.800-218

[30] Hazell, J. (2023). *Spear Phishing with Large Language Models*. arXiv. https://doi.org/10.48550/arXiv.2305.06972

[31] The MITRE Corporation (2025). MITRE ATT&CK®. https://attack.mitre.org/

[32] Hackett, W; Birch, L; Trawicki, S; Suri, N; & Garraghan, P (2025). *Bypassing LLM Guardrails: An Empirical Analysis of Evasion Attacks against Prompt Injection and Jailbreak Detection Systems.* arXiv. https://doi.org/10.48550/arXiv.2504.11168

[33] Plesner, A; Vontobel, T; & Wattenhofer, R (2024). *Breaking reCAPTCHAv2*. arXiv. https://doi.org/10.48550/arXiv.2409.08831

[34] Dhanushkodi, K; Thejas, S (2024). *AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation.* https://doi.org/10.1109/ACCESS.2024.3493957

[35] Rose, S; Borchert, O; Mitchell, S; & Connelly, S (2020). Zero Trust Architecture (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-207. https://doi.org/10.6028/NIST.SP.800-207

[36] Huang, L. et al. (2024). *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions.* arXiv. https://doi.org/10.48550/arXiv.2311.05232

[37] Ponomareva, N. et al. (2023). *How to DP-fy ML: A Practical Guide to Machine Learning with Differential Privacy*. arXiv. https://doi.org/10.48550/arXiv.2303.00654

[38] Abdali, S; Anarfi, R; Barberan, CJ; & Shayegani, E (2025). *Securing Large Language Models: Threats, Vulnerabilities and Responsible Practices*. arXiv. https://doi.org/10.48550/arXiv.2403.12503

[39] Girhepuje, S; Verma, A; & Raina, G (2024). *A Survey on Offensive AI Within Cybersecurity*. arXiv. https://doi.org/10.48550/arXiv.2410.03566

[40] Chen, Z; Ji, C (2005.) *A self-learning worm using importance scanning.* https://doi.org/10.1145/1103626.1103632

[41] Avgerinos, T. et al. (2014). *Automatic exploit generation.* https://doi.org/10.1145/2560217.2560219

[42] National Institute of Standards and Technology (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (National Institute of Standards and Technology, Gaithersburg, MD), NIST AI 600-1. https://doi.org/10.6028/NIST.AI.600-1

[43] Castagnaro, A; Conti, M; & Pajola, L (2024). *Offensive AI: Enhancing Directory Brute-forcing Attack with the Use of Language Models.* arXiv. https://doi.org/10.48550/arXiv.2404.14138

[44] Kure, H. et al. (2025). *Detecting and Preventing Data Poisoning Attacks on AI Models.* arXiv. https://doi.org/10.48550/arXiv.2503.09302
[45] Souly, A. et al. (2025). *Poisoning Attacks on LLMs Require a Near-constant Number of Poison Samples*. arXiv. https://doi.org/10.48550/arXiv.2510.07192
[46] Wipro (2025). State of Cybersecurity Report 2025. https://www.wipro.com/cybersecurity/reports/state-of-cybersecurity-report-2025/
[47] Booth, H et. al. D (2024). Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-218A. https://doi.org/10.6028/NIST.SP.800-218A
[48] Google Threat Intelligence Group (2025). *GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools*. https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools
[49] Microsoft (2025). Microsoft Digital Defense Report 2025. https://www.microsoft.com/en-us/corporate-responsibility/cybersecurity/microsoft-digital-defense-report-2025/
[50] Open Worldwide Application Security Project (2025). Generative AI and Agentic Security Solutions Landscape. https://genai.owasp.org/resource/owasp-genai-security-project-solutions-reference-guide-q2_q325/
[51] National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (RMF) Playbook (National Institute of Standards and Technology, Gaithersburg, MD). https://airc.nist.gov/airmf-resources/playbook/
[52] Cloud Security Alliance (2025). AI Controls Matrix. https://cloudsecurityalliance.org/artifacts/ai-controls-matrix

# Appendix A. List of Symbols, Abbreviations, and Acronyms

**AI**
Artificial Intelligence

**AIBOM**
Artificial Intelligence Bill of Materials

**AI RMF**
Artificial Intelligence Risk Management Framework

**API**
Application Programming Interface

**APT**
Advanced Persistent Threat

**BDA**
Big Data Analytics

**CISO**
Chief Information Security Officer

**CMMC**
Cybersecurity Maturity Model Certification

**CMMI**
Capability Maturity Model Integration

**COI**
Community of Interest

**COSAIS**
Control Overlays for Securing AI Systems

**CPS**
Cyber-Physical System

**CPRT**
Cybersecurity and Privacy Reference Tool

**CSF**
Cybersecurity Framework

**CTI**
Cyber Threat Intelligence

**DDoS**
Distributed Denial-of-Service

**DFARS**
Defense Federal Acquisition Regulation Supplement

**EOL**
End-of-Life

**FedRAMP**
Federal Risk and Authorization Management Program

**FN**
False Negative

**FP**

818 False Positive

819 **GenAI**
820 Generative Artificial Intelligence

821 **GPU**
822 Graphics Processing Unit

823 **HITL**
824 Human-in-the-loop

825 **ICS**
826 Industrial Control Systems

827 **IDS**
828 Intrusion Detection Systems

829 **IoT**
830 Internet of Things

831 **IR**
832 Informative Reference

833 **ISAC**
834 Information Sharing and Analysis Center

835 **ISAO**
836 Information Sharing and Analysis Organization

837 **ITL**
838 Information Technology Laboratory

839 **LLM**
840 Large Language Model

841 **MFA**
842 Multi-Factor Authentication

843 **ML**
844 Machine Learning

845 **NCCoE**
846 National Cybersecurity Center of Excellence

847 **NIST**
848 National Institute of Standards and Technology

849 **OpenCTI**
850 Open Cyber Threat Intelligence

851 **OSINT**
852 Open Source Intelligence

853 **PII**
854 Personally Identifiable Information

855 **RL**
856 Reinforcement Learning

857 **RMF**
858 Risk Management Framework

859 **SaaS**

Software as a Service

**SARSA**
State-Action-Reward-State-Action

**SBOM**
Software Bill of Materials

**SP**
Special Publication

**SQL**
Structured Query Language

**STIX**
Structured Threat Information eXpression

**TPU**
Tensor Processing Unit

**UEBA**
User and Entity Behavior Analytics

# Appendix B. Glossary

This appendix will include terms defined and explained to facilitate a reader's comprehension of the report.

## Appendix C. Cybersecurity Framework 2.0 Overview

This Community Profile is based on the NIST Cybersecurity Framework (CSF) 2.0, which provides a flexible and risk-based approach to managing cybersecurity. The CSF helps organizations of any size, sector, or level of cyber maturity address their unique cybersecurity risks while improving communication and collaboration across stakeholders.

The CSF Core is a taxonomy of high-level cybersecurity outcomes that can help organizations manage their cybersecurity risks. The Core components are a hierarchy of Functions, Categories, and Subcategories that detail each outcome. These outcomes can be understood by a broad audience, including executives, managers, and practitioners, regardless of their cybersecurity expertise.

The CSF Core Functions (Govern, Identify, Protect, Detect, Respond, Recover) organize cybersecurity outcomes at their highest level. The Core then identifies underlying Categories and Subcategories for each Function. The six Functions of the CSF 2.0 Core are composed of 22 Categories, which are further broken down into 106 Subcategories of more specific outcomes. The Core structure is depicted in Fig. 3 below.

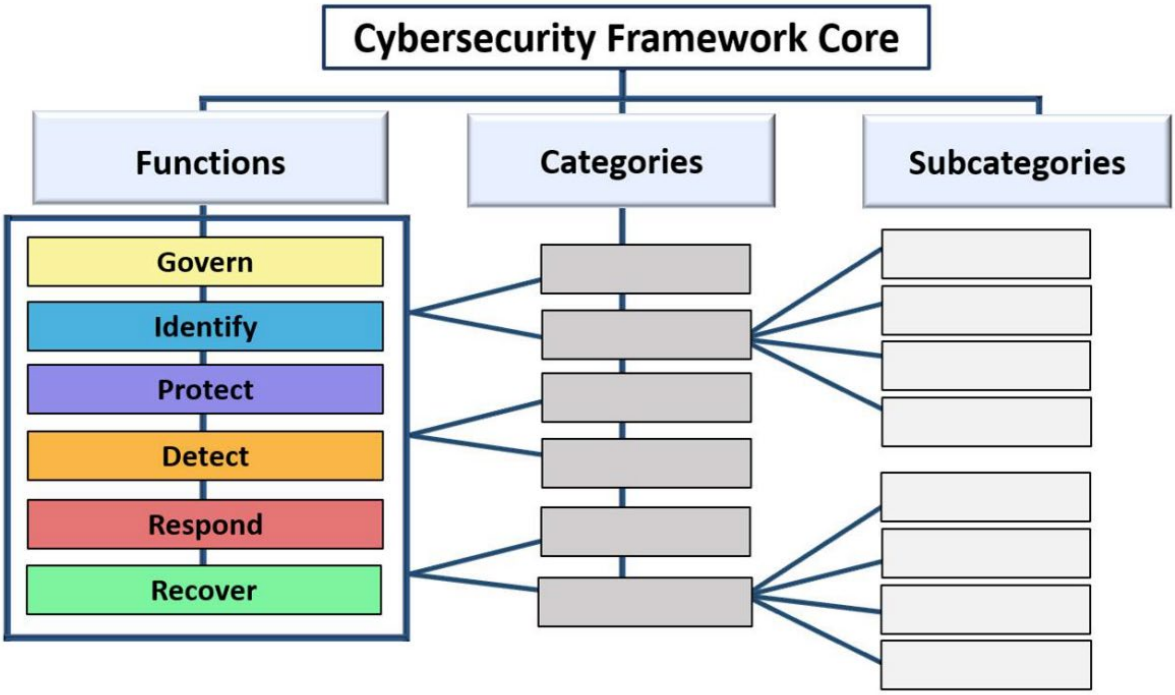



**Fig. 3: CSF Core Structure**

Sections 2.3-2.8 contain the descriptions of the CSF Functions, Categories, and Subcategories that are relevant to the Cyber AI Profile. Additionally, the NIST CSF website provides supplemental resources to help organizations understand, adopt, and use the CSF and achieve the desired outcome of each Subcategory, including Implementation Examples and Informative References. Communities and organizations can choose to add their own Implementation Examples and Informative References that are unique to their environment in the future.

## Appendix D. How to Use the Cyber AI Profile

The Cyber AI Profile is designed to support an organization's cybersecurity program with understanding cybersecurity-related opportunities for implementing AI capabilities and managing new cybersecurity risks introduced by AI systems, regardless of where they are on their AI journey. Use of the Cyber AI Profile will:

- Provide organizations with proposed priorities for achieving the outcomes described in of the CSF Core;
- Assist organizations in understanding how AI advances can enhance approaches to cybersecurity risk management;
- Assist organizations in understanding where AI advances may introduce new cybersecurity risks; and
- Provide a common approach for discussing the nexus of cybersecurity and AI.

The Cyber AI Profile provides guidance that can be adapted based on an organization's maturity level, available resources, needs, and goals. For example, an organization may want to prioritize, remove, or add Focus Areas. An organization may also adjust the proposed priorities of the Subcategories, or choose a different prioritization schema.

The Cyber AI Profile provides an informed starting point for organizations to identify and address desired cybersecurity outcomes for AI. An organization can use the information in the Cyber AI Profile to support risk management activities, such as:

- Benchmarking internal progress against community priorities and considerations (e.g., when creating or updating a Current Profile);
- Informing Target Profile(s);
- Communicating the significance of impacts when analyzing the gaps between Current and Target Profiles;
- Facilitating decision making when allocating budget, personnel, and other resources;
- Determining which cybersecurity outcomes are most important to assess and validate as part of the organization's cybersecurity risk management program;
- Developing implementation and action plans; and
- Communicating expectations with vendors, partners, and other external stakeholders.

While the Cyber AI Profile provides insights to help organizations make decisions, it is up to each organization to determine its priorities and how to achieve the outcomes in the CSF Core. For example, organizations may choose to address one Focus Area at a time, instead of all three

932 at once. Organizations may also make decisions like addressing only the High Priority (1)
933 Subcategories first, before addressing the others.