The CRI Financial Services   
AI Risk Management Framework   
Version 1.0

Guidebook – Initial AI Adoption Stage

February 2026

**

Developed in collaboration with the Financial Services Sector Coordinating Council and the U.S. Department of the Treasury

Cyber Risk Institute  
Washington, D.C.

[CyberRiskInstitute.org](http://www.cyberriskinstitute.org/)

**ABOUT THE CYBER RISK INSTITUTE**

The Cyber Risk Institute (CRI) is a non-profit coalition of financial institutions and trade associations dedicated to advancing the development and harmonization of cybersecurity, technology, and artificial intelligence (AI) risk management standards. CRI is committed to providing practical, industry-driven solutions for emerging risks, including AI, and leverages its expertise in cyber and technology risk management as a solid foundation for AI governance and risk management.

By providing clear guidance and support, CRI assists organizations in mitigating new and emerging threats while enhancing risk practices across the financial sector. Building on CRI's existing work in standardization and simplification, including the Profile and Cloud Profile[[1]](#footnote-2), CRI is proud to deliver the **Financial Services Artificial Intelligence Risk Management Framework** (hereinafter referred to as the “FS AI RMF” or “the Framework”)[[2]](#footnote-3).

**ABOUT THE FS AI RMF**

The FS AI RMF was developed through public-private collaboration involving a diverse group of 108 financial institutions, including community banks, credit unions, national and multinational banks, investment firms, insurance companies, and trade associations under the Financial Services Sector Coordinating Council (FSSCC). Additionally, U.S. and international agencies, most notably the National Institute of Standards and Technology (NIST), provided input and feedback. This industry-backed, consolidated approach is designed to help organizations of all types and sizes effectively identify, evaluate, manage, and govern AI risks. It aims to operationalize a comprehensive risk management framework tailored specifically for financial services, promoting efficiency and innovation while harnessing the potential of AI both strategically and responsibly.

**Alignment with Existing Risk Management Functions**

The FS AI RMF is designed to be a universal complement to an organization’s existing risk management policies and practices. It does not replace existing frameworks, serve as an all-inclusive repository of risks and controls, or prescribe specific use-case or role-specific guidance. That said, CRI recognizes AI as an enterprise risk. Therefore, the Framework is designed to logically connect to broader enterprise governance, risk, and compliance programs, supporting integrated risk assessments[[3]](#footnote-4), aggregation, prioritization, and mitigation efforts.

**Alignment with Existing Frameworks, Standards, and Regulatory Guidance**

Structurally aligned with the NIST AI Risk Management Framework (hereinafter referred to as the “NIST AI RMF”), the FS AI RMF synthesizes insights from multiple global regulations, standards, official guidance, and supervisory provisions.[[4]](#footnote-5) It expands upon NIST’s pre-existing Functions, Categories, and Sub-Categories through 230 Control Objectives that support organizations at various stages of AI adoption. Rather than replacing or restructuring current standards, it extends them within a financial sector context, facilitating harmonization of standards and regulations globally.

Designed to be adaptable and scalable, the FS AI RMF incorporates clear risk definitions, staged AI adoption levels, and comprehensive Control Objectives. This approach enables financial institutions to harness AI’s transformative potential responsibly—driving industry progress while safeguarding the integrity of the financial ecosystem.

**ABSTRACT**

This Guidebook provides financial services organizations with comprehensive guidance for implementing CRI’s FS AI RMF. It serves as a helpful resource that explains how to effectively navigate and utilize the AI Adoption Stage Questionnaire and Risk and Control Matrix (RCM), offering detailed recommendations and contextual information for a firm’s risk identification, assessment, and mitigation practices related to AI adoption and use. The Guidebook reflects sector-specific considerations and aligns global standards, guidelines, and regulatory expectations. It is intended to support consistent, responsible AI practices within financial institutions’ existing governance, risk, and compliance frameworks.

**ACKNOWLEDGEMENTS**

The FS AI RMF reflects the dedicated efforts of more than 100 organizations. CRI sincerely thanks the financial institutions, trade associations, and other stakeholders who contributed to this important initiative. Special gratitude goes to KPMG for their ongoing support, manpower, and advisory throughout the Framework’s development, as well as to NIST and HallResearch.ai for expert technical advice and continuous AI thought leadership.

We also extend our thanks to our 150+ member organizations for supporting CRI initiatives, the CRI Board of Directors for strategic guidance, and the 100+ active participants of the FS AI RMF Working Group for reviewing and validating the Framework’s content. Finally, we would like to acknowledge CRI’s Angela Patel, John Goodman, and Emily Beam for their leadership, oversight, and engagement.

**CONTENTS**

[UPDATE SCHEDULE AND VERSIONS 5](#_Toc221703736)

[INTRODUCTION 6](#_Toc221703737)

[Background and Motivation 6](#_Toc221703738)

[Purpose and Benefits 6](#_Toc221703739)

[Audience 6](#_Toc221703740)

[Scope 7](#_Toc221703741)

[FS AI RMF COMPONENTS 7](#_Toc221703742)

[Adoption Stage Questionnaire 7](#_Toc221703743)

[Risk and Control Matrix 8](#_Toc221703744)

[Guidebook 10](#_Toc221703745)

[Control Objective Reference Guide 10](#_Toc221703746)

[SECTION 1: THE AI ADOPTION STAGE QUESTIONNAIRE 11](#_Toc221703747)

[Overview 11](#_Toc221703748)

[Instructions 12](#_Toc221703749)

[The AI Adoption Stage Questionnaire 12](#_Toc221703750)

[SECTION 2: THE RISK AND CONTROL MATRIX 15](#_Toc221703751)

[Overview 15](#_Toc221703752)

[Control Objective Names and Descriptions 16](#_Toc221703753)

[APPENDIX A – AI TRUSTWORTHY PRINCIPLES 21](#_Toc221703754)

[APPENDIX B – ACRONYMS AND ABBREVIATIONS 22](#_Toc221703755)

[APPENDIX C – GLOSSARY OF KEY TERMS 24](#_Toc221703756)

[APPENDIX D – INFORMATIVE REFERENCES 29](#_Toc221703757)

[APPENDIX E – ADDITIONAL SOURCE MATERIALS 31](#_Toc221703758)

# **UPDATE SCHEDULE AND VERSIONS**

The FS AI RMF and this Guidebook are living documents. CRI will review the content and usefulness regularly to determine if an update is or may become necessary. The nature and extent of the update will be reviewed with the CRI Board of Directors, members, stakeholders, trade associations, relevant third parties, and others where appropriate.

Updates will be tracked in a Version Control Log using a consistent, two-number versioning schedule. The Version Control Log will include: (1) the version number; (2) the date of the change(s); (3) a description of the change(s); (4) the updated section(s); and (5) the relevant page number(s). For example:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date of Change(s)** | **Description of Change(s)** | **Relevant Section(s)** | **Page Number(s)** |
| 1.0 | XX/XX/XXX | [Change] | [Section] |  |

General comments and change requests should be sent to [Secretariat@cyberriskinstitute.org](mailto:Secretariat@cyberriskinstitute.org) and will be reviewed at least semi-annually.

# **INTRODUCTION**

Background and Motivation

As AI adoption in the financial sector continues to accelerate, it introduces significant opportunities as well as risks—such as bias, opacity, cybersecurity vulnerabilities, and systemic interdependencies—that existing frameworks are not designed to address. Many organizations believe they align with the principles outlined in the 2023 NIST AI RMF, but applying these principles to advanced AI systems, like large language models (LLMs), reveals gaps due to complexity and novelty[[5]](#footnote-6).

Experts emphasize that evaluating, benchmarking, and managing risks tied to such sophisticated models demand more practical, targeted guidance beyond traditional risk frameworks. These models often pose unique challenges that include ensuring fairness, explainability, and cybersecurity resilience, which are difficult to assess with current standards based on conventional technology.

Recognizing this gap, the industry seeks a tailored AI risk management approach—one that evolves with AI technological advances and supports financial institutions in deploying AI responsibly, securely, and transparently. Developing such a framework aims to strengthen governance, foster trust, mitigate risks, and ensure that AI-driven innovation benefits the financial ecosystem without introducing undue vulnerabilities.

Purpose and Benefits

In response to industry recognition of the limitations of existing frameworks and the need for targeted guidance, the FS AI RMF Guidebook aims to: (1) help organizations implement the FS AI RMF effectively, and (2) promote sector-wide adoption and understanding of the Framework. It provides detailed guidance on the 230 Control Objectives described in the RCM, aligned with the four stages of AI adoption defined in the AI Adoption Stage Questionnaire.

Building on ongoing efforts to address AI risks, this Guidebook supports consistent evaluation, benchmarking, and maturity in AI governance. It offers practical guidance for developing, implementing, and maturing AI risk and control programs, fostering trust among stakeholders, and demonstrating a commitment to responsible AI deployment. Ultimately, it enables organizations to better manage risks and capitalize on AI benefits to build a resilient, innovative financial services industry.

Audience

The FS AI RMF is designed for all financial institutions, regardless of size, type, complexity, or criticality. It provides a structured approach that both organizations and their third-party providers can use to evaluate and manage AI-related risks throughout the supply chain. While the Framework does not assign specific roles to particular objectives or activities, many types of users will find the Guidebook valuable. Relevant roles include AI professionals (e.g., engineers, data scientists, architects, developers, researchers), enterprise technology leaders, risk and compliance managers, and legal professionals.

Individuals with a strong understanding of financial markets, regulations, products, and AI/ML technologies are also encouraged to engage with this guidance. The primary audience for the Guidebook includes organizations seeking to assess their current AI adoption level, define their desired future state, and implement the necessary Control Objectives to bridge the gap effectively. It is assumed that readers possess baseline knowledge of AI and AI risk management concepts to understand and apply the Framework’s components effectively.

Scope

The scope of this Guidebook is to provide guidance for firms in Adoption Stage 1 – Initial – to implement the relevant Control Objectives outlined within the FS AI RMF. The Guidebook’s companion document, “Control Objective Reference Guide – Initial Stage,” offers additional details, including examples of Controls and Effective Evidence. Both documents support organizations in understanding how to interpret and apply this information at various stages of AI adoption, from initial AI deployment to deeply embedded, enterprise-wide AI implementation.

# **FS AI RMF COMPONENTS**

The FS AI RMF has four primary components:

1. An **AI** **Adoption Stage Questionnaire** **(the “Questionnaire”)** which helps organizations better understand their current state of AI adoption;
2. A **Risk and Control Matrix (RCM)**, comprised of Risk Statements and Control Objectives and provides practical implementation guidance organized by AI Adoption Stage;
3. This **Guidebook**, which describes how to leverage and deploy the FS AI RMF; and
4. The **Control Objective Reference Guide**, which offers additional details for each Control Objective and lists examples of corresponding Controls and Effective Evidence.

Adoption Stage Questionnaire

The Questionnaire guides organizations through a self-assessment that helps determine their current AI Adoption Stage, classifying them into one of the four stages defined below:

1. **Initial** features limited and protective methods;
2. **Minimal** entails low-risk implementations;
3. **Evolving** involves high-risk production applications; and
4. **Embedded** describes AI that is widely integrated across the organization.

These four stages consider the business impact of AI, technology implementation, and scalability.

|  | **Initial** | **Minimal** | **Evolving** | **Embedded** |
| --- | --- | --- | --- | --- |
| **Business Impact of AI** | AI is not embedded in critical functions or business decisions. | AI use is limited (production or non-production) for non-critical tasks. | AI drives outcomes through external-facing solutions and the use of sensitive data, but it is not utilized for critical decision-making. | AI drives outcomes by incorporating autonomous decision-making and data-driven insights into critical business functions. |
| **Technology Implementation** | Predictive AI and model risk management is used but relies on legacy systems with no adoption of modern AI technologies, external-facing solutions, or internal model development. | AI is narrowly in use and is not external-facing. AI does not handle sensitive data and there is no internal model development. | AI adoption includes expanded use of solutions and sensitive data processing but lacks internal model development. | AI is used for autonomous decision-making, high-sensitivity data processing, external-facing solutions, internal model development. |
| **Scalability** | Existing AI systems lack scalability and are confined to specific, limited applications that do not interact with or support broader organizational processes. | AI applications are limited to isolated instances without integration into wider business processes, and the organization lacks the infrastructure necessary to scale AI solutions across multiple areas or functions. | AI systems show potential for wider application, with deployment in more extensive, external-facing functions, but these systems are not fully scalable or integrated, indicating a transition towards broader usability. | AI systems are fully scalable and integrated throughout the organization. A robust framework supports the deployment and interconnection of AI across diverse functions, promoting growth and adaptability. |

To determine its appropriate AI Adoption Stage, the AI Questionnaire prompts the user to evaluate six key dimensions—Business Impact, Governance, Deployment Model, Third-Party AI Use, Organizational Goals, and Data Sensitivity and Criticality—by reviewing statements that describe their current practices and policies. Based on the responses, organizations will determine whether they are in the early, mid, or later stages of AI adoption. This provides the foundation for understanding the organization’s current and desired AI use and enables a more customized view of the related Risk Statements and applicable Control Objectives defined in detail in the RCM.

Please see **Section 1** for the Adoption Stage Questionnaire.

**Important Note**: Section 2 of this Guidebook pertains only to the Control Objective information relevant to the Stage 1 - Initial. It should only be used by organizations at the "Initial" stage of adoption, or by organizations looking for a "quick start" guide focused only on the most fundamental Control Objectives. For Control Objectives related to the other three AI Adoption Stages, please see the Full Guidebook.

Risk and Control Matrix

The RCM provides a structured, comprehensive way to organize Control Objectives by Adoption Stage and align them with AI Trustworthy Principles and important risk information. It consists of the information outlined below, which is further detailed in **Section 2**.

|  |  |
| --- | --- |
| **NIST Function, Category, and Sub-Category** | The structure of the RCM is based on the Functions, Categories, and Sub-Categories defined in the NIST AI RMF. Each includes a Reference ID, Name, and Description. Control Objectives are mapped to the relevant Sub-Category, which then rolls up into a Category and finally into a Function. This classification offers a standardized taxonomy to organize risk and control information, facilitating consistency and clarity in risk management practices. |
| **AI Trustworthy Principle and Risk Statement** | Each Control Objective is aligned with its primary AI Trustworthy Principle (e.g., Accountable, Transparent, Explainable, Fair) and a broad Risk Statement. This is done for the purposes of prioritization and aggregation within the Framework. In practice, several Trustworthy Principles may apply, and Risk Statements may be customized to connect to an firm’s risk taxonomy. |
| **Control Objectives** | These are specific policies, processes, or operational actions within each Sub-Category that should be implemented to mitigate AI-related risks. Each includes a Reference ID, Name, and Description. Control Objectives cover a broad range of topics such as data quality, fairness, security, transparency, and operational resilience. Each Control Objective is designed not to be prescriptive but to promote practical Control implementation suited to current AI Adoption Stage and organizational needs. |
| **Implementation Guidance** | This provides actionable and practical recommendations on how best to implement each Control Objective. While the Implementation Guidance is not meant to be an exhaustive set of instructions, it should point an organization in the right direction when determining how Control Objectives should be put into practice in their unique environments. |

As mentioned, Control Objectives are also aligned with the Adoption Stage(s) to which they apply. For example, certain Control Objectives might only apply to organizations at the Evolving or Embedded stage. This helps organization at earlier stages in particular, as it allows them to focus on a smaller number of key Control Objectives as opposed to using limited resources for Control Objectives that may not be relevant yet.

**Navigating the RCM**

The RCM is designed to be flexible and adaptable, and as such, there are a number of ways to look at the information contained within it. The general structure is based on NIST’s Functions, Categories, and Sub-Categories, with CRI-defined Control Objectives defined at the Sub-Category level. There will typically be 3-5 control objectives per Sub-Category; however, in some cases there will be more or less. Each of these components has an ID, a Name, and a Description. For example:

Function ID: GV

Function Name: Govern

Function Description: A culture of risk management is cultivated and present.

Category ID: GV-1

Category Name: Establishing Key Policies & Processes

Category Description: Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively.

Sub-Category ID: GV-1.1

Sub-Category Name: Legal and Regulatory Compliance

Sub-Category Description: Legal and regulatory requirements involving AI are understood, managed, and documented.

Control Objective ID: GV-1.1.1

Control Objective Name: AI Legal, Regulatory, and Policy Integration

Control Objective Description: The organization identifies, monitors, and integrates applicable laws, regulations, contractual obligations, and sector requirements into policies, procedures, and operations governing AI. This includes updating governance artifacts, aligning operational practices, and verifying ongoing compliance as requirements evolve.

A more detailed overview of the RCM’s structure as well as navigational tips can be found in **Section 2**.

Guidebook

The Guidebook provides practical step-by-step guidance to support FS AI RMF implementation. In addition to the background information provided above, this Guidebook includes:

1. **Section 1**: The Adoption Stage Questionnaire and instructions for completing it
2. **Section 2**: An overview of the RCM and its contents
3. **Appendix A**: NIST AI Trustworthy Principles
4. **Appendix B**: Acronyms and Abbreviations
5. **Appendix C**: Glossary of Key Terms
6. **Appendix D**: Informative References
7. **Appendix E**: Additional Source Materials

Control Objective Reference Guide

The Control Objective Reference Guide serves as a companion to this Guidebook. It encompasses all the information contained in Section 2 of this document, along with five illustrative example Controls for each Control Objective and four examples of Effective Evidence for each Control.

**IMPORTANT:** Please note that the examples of Controls and Effective Evidence provided in the Control Objective Reference Guide are intended to serve only as illustrative guides. They are not exhaustive, exclusive, or all-encompassing lists and should not be interpreted as a definitive guarantee for meeting regulatory or audit expectations. The suitability and applicability of these examples may vary across organizations, and different Evidence might provide greater value in specific contexts.

Users of the FS AI RMF should critically assess the adequacy and relevance of the Controls and Evidence presented to adequately substantiate their responses, weighing it against the resources and investment required to produce such Evidence. It is important to note that there should be no presumption that all, or even most, of the examples of Controls and Effective Evidence included are necessary, advisable, or appropriate for supporting the organization’s compliance assertions. Organizations are encouraged to tailor their approach to their specific context and regulatory environment.

# **SECTION 1: THE AI ADOPTION STAGE QUESTIONNAIRE**

**Important Note**: This Guidebook pertains only to the Control Objective information relevant to the Stage 1 - Initial. It should only be used by organizations at the "Initial" stage of adoption, as determined by this Questionnaire, or by organizations looking for a "quick start" guide focused on only the most fundamental Control Objectives. After completing this Questionnaire, if your organization is at Adoption Stages 2, 3, or 4, please see the Full Guidebook.

Overview

The Questionnaire guides organizations through a self-assessment that helps determine their current AI Adoption Stage, classifying it into one of the four stages defined below:

1. **Initial** features limited and protective methods;
2. **Minimal** entails low-risk implementations;
3. **Evolving** involves high-risk production applications; and
4. **Embedded** describes AI that is widely integrated across the organization.

Each of these four stages consider:

1. Business Impact of AI;
2. Technology Implementation; and
3. Scalability.

**Stage 1: Initial**

|  |  |
| --- | --- |
| **Business Impact** | AI is not embedded in critical functions or business decisions. |
| **Technology** | Utilizes predictive AI and employs model risk management, relying on legacy AI systems with no adoption of modern AI technologies, external facing solutions, or internal model development. |
| **Scalability** | Existing AI systems lack scalability and are confined to specific, limited applications that do not interact with or support broader organizational processes. |

**Stage 2: Minimal**

|  |  |
| --- | --- |
| **Business Impact** | Limited use of AI (production or non-production), used for non-critical tasks. |
| **Technology** | AI is narrowly in use and is non-external facing. AI does not handle sensitive data and there is no internal model development. |
| **Scalability** | AI applications are limited to isolated instances without integration into wider business processes. Organization lacks the infrastructure necessary to scale AI solutions across multiple areas or functions. |

**Stage 3: Evolving**

|  |  |
| --- | --- |
| **Business Impact** | AI drives outcomes through external-facing solutions and use of sensitive data. Not utilized for critical business decision-making. |
| **Technology** | AI adoption includes expanded use of solutions and sensitive data processing but lacks internal model development. |
| **Scalability** | AI systems begin to show potential for wider application, with deployment in more extensive, external-facing functions. However, these systems are not yet fully scalable or integrated across the entire enterprise, indicating a transitional phase towards broader usability. |

**Stage 4: Embedded**

|  |  |
| --- | --- |
| **Business Impact** | AI drives outcomes by incorporating autonomous decision-making and data-driven insights into critical business functions. |
| **Technology** | Use of autonomous decision-making, high-sensitivity data processing, external facing solutions, internal model development. |
| **Scalability** | AI systems are fully scalable and integrated throughout the organization. There is a robust framework that supports the deployment and interconnection of AI across diverse organizational functions, promoting comprehensive growth and adaptability. |

Instructions

In order to determine your organization's AI Adoption Stage, please complete the steps below.

Each step will prompt you to evaluate six statements and determine if any of these statements reflect your organization. Each set of statements corresponds to different organizational dimensions, presented in a randomized order.

1. Business Impact
2. Governance
3. Deployment Model
4. Third-Party AI Use
5. Organizational Goals
6. Data Sensitivity and Criticality

Starting with Step 1, review the six statements to determine if, to the best of your knowledge, you believe one or more of those statements represents your organization's stage of AI Adoption.

1. If at least one of the six statements aligns with your organization, select "Yes,” and stop. That is your organization’s current AI Adoption Stage.
2. If none of the statements aligns with your organization, select “No” and continue on to the next step; and repeat this exercise until you answer "Yes."

**Results**: A "Yes" answer marks the conclusion of the Questionnaire and provides your organization's AI Adoption Stage.

Note that the AI Adoption Stage Questionnaire and these instructions can also be found in the first 4 tabs of the RCM workbook.

The AI Adoption Stage Questionnaire

**STEP 1**

|  |
| --- |
| 1. In my organization, AI systems are deeply integrated into critical business processes, transforming operations, and regularly handling sensitive and/or regulated data. |
| 2. My organization has an extensive, enterprise-wide, AI-enabled governance program overseeing the use of AI systems, regularly assessed to align with risk and regulatory compliance requirements. |
| 3. My organization deploys AI systems through a combination of vendor-managed platforms, internally hosted models, and custom-built solutions, fully integrated across the organization. |
| 4. My organization extensively relies on third-party AI vendors for a wide range of functions and use cases, including chatbots, risk models, and other critical applications, with deep integration into internal systems and processes. |
| 5. My organization's key goals in adopting AI are strategic differentiation, innovation, scalability, transformative business models, and ethical business practices, in addition to operational efficiency, customer experience, safety, and regulatory compliance. |
| 6. My organization has deeply integrated AI into critical business functions, driving autonomous decision-making and data-driven insights, with enterprise-wide adoption. |

If you answered “Yes” to any of these statements, **stop**. Your organization’s level of adoption is **Stage 4: Embedded**. For information about the Control Objectives corresponding to this adoption stage, please refer to the **“Risk and Control Matrix – Embedded Stage”** workbook.

If you answered “No,” continue to Step 2.

**STEP 2**

|  |
| --- |
| 1. My organization uses AI systems in key business processes, in some cases transforming operations and/or processing sensitive or regulated data types in specific areas. |
| 2. My organization has governance policies for AI systems, but they are not consistently applied across the entire organization. There are ongoing efforts to align governance and risk policies to regulatory requirements, but coverage is not yet all-inclusive. |
| 3. My organization deploys AI systems through a mix of vendor-managed platforms and internally hosted models, with some integration across specific functions. |
| 4. My organization relies on third-party AI vendors for specific functions or use cases, with some integration into internal systems and processes. |
| 5. My organization's key goals driving AI adoption are improving operational efficiency, enhancing customer experience, and ensuring regulatory compliance, with some focus on innovation and new business opportunities. |
| 6. In my organization, AI supports external-facing outcomes but is not used for critical business decision-making and is limited to specific functions. |

If you answered “Yes” to any of these statements, **stop**. Your organization’s level of adoption is **Stage 3: Evolving**. For information about the Control Objectives corresponding to this adoption stage, please refer to the **“Risk and Control Matrix – Evolving Stage”** workbook.

If you answered “No,” continue to Step 3.

**STEP 3**

|  |
| --- |
| 1. My organization does not use AI systems in key business processes, for operational transformation, and/or to process sensitive or regulated data types. |
| 2. My organization has limited to no governance policies specific to AI systems, and oversight is handled on a case-by-case basis. |
| 3. In my organization, AI systems are primarily deployed through vendor-managed platforms, with limited internal hosting or integration. |
| 4. My organization has limited reliance on third-party AI vendors, with minimal integration into internal systems and processes. |
| 5. The key organizational goals driving AI adoption in my organization are primarily focused on improving operational efficiency and reducing costs. |
| 6. In my organization, AI has limited use in non-critical tasks and is confined to specific use cases. |

If you answered “Yes” to any of these statements, **stop**. Your organization’s level of adoption is **Stage 2: Minimal**. For information about the Control Objectives corresponding to this adoption stage, please refer to the **“Risk and Control Matrix – Minimal Stage”** workbook.

If you answered “No,” continue to Step 4.

**STEP 4**

|  |
| --- |
| 1. My organization is not currently using AI systems in any processes but is exploring potential use cases that may involve critical processes or sensitive data types. |
| 2. My organization does not have governance policies for AI systems but is exploring potential frameworks and guidelines. |
| 3. My organization is not currently deploying any AI systems but is exploring vendor-managed platforms or internally hosted models for specific use cases. |
| 4. My organization does not currently rely on third-party AI vendors but is exploring potential partnerships or solutions for specific functions or use cases. |
| 5. My organization is exploring the potential for AI to improve operational efficiency, customer experience, and regulatory compliance but has not yet defined clear goals or priorities. |
| 6. In my organization, AI is not currently used in production or non-production environments, and its potential is being explored for specific use cases. |

If you answered “Yes” to any of these statements, **stop**. Your organization’s level of adoption is **Stage 1: Initial**. For information about the Control Objectives corresponding to this adoption stage, please refer to the **“Risk and Control Matrix – Initial Stage”** workbook.

If you answered “No,” continue to Step 5.

**STEP 5**

|  |
| --- |
| 1. My organization does not currently use AI systems in any processes and does not process sensitive or regulated data types. |
| 2. My organization does not currently have governance policies for AI systems. |
| 3. My organization does not currently deploy AI systems, nor does it plan to in the near future. |
| 4. My organization does not currently rely on third-party AI vendors. |
| 5. My organization has not identified any specific goals for AI adoption and is not actively considering AI as a strategic priority. |
| 6. In my organization, AI is not currently used to support any business functions. |

If you answered “Yes” to any of these statements, your organization has yet to adopt AI.

# **SECTION 2: THE RISK AND CONTROL MATRIX**

Once you have determined your organization’s current AI Adoption Stage, you can customize the view of the RCM by filtering only on your current and/or desired level of adoption. Firms who are in the “Initial” stage, for example, will not need to review all the same Control Objectives as firms in the “Embedded” stage. This feature of the RCM makes it easy to prioritize what is most important and relevant for your organization.

Overview

**Important Note:** This Guidebook pertains only to the Control Objective information relevant to the Stage 1 - Initial. It should only be used by organizations at the "Initial" stage of adoption, or by organizations looking for a "quick start" guide focused only on the most fundamental Control Objectives. For Control Objectives related to one of the other three AI Adoption Stages, please see the Full Guidebook.

**Structure:** The FS AI RMF, like the NIST AI RMF, includes four overarching Functions for evaluating an organization’s AI risk management program: (1) Govern; (2) Map; (3) Measure; and (4) Manage. Each of these Functions is sub-divided into specific concept Categories and Subcategories, which are designed to reflect an element of an effective AI Framework.

**Control Objectives**: Each Subcategory is associated with at least one Control Objective. A list of all 21 Initial Stage Control Objective Names and Descriptions can be found in the tab marked “View 1” in the RCM workbook and in the “Risk and Control Matrix Details” section beginning on the next page.

**Relevant AI Adoption Stage(s)**: After completing the AI Adoption Questionnaire and determining your AI Adoption Stage - Initial - you can focus on the 21 Control Objectives corresponding to that AI Adoption Stage. A list of all 21 Control Objectives for the Initial Stage, along with a justification for the applicability of each Control Objective to this Adoption Stage, is also included in “View 2-All-Initial” found in the RCM.

**Corresponding Risk Names, Risk Statements, and AI Trustworthy Principles**: In addition to the Control Objectives and Adoption Stage information, the "View 2-All-Initial" tab of the RCM workbook also contains other important information. Each Control Objective has a corresponding Risk Name and Risk Statement and is aligned to a primary AI Trustworthy Principle. Practical Implementation Guidance is also provided

**Sample Controls and Examples of Effective Evidence**: Five examples of possible Controls and 4 examples of Effective Evidence per Control have been documented for each Control Objective. This information is illustrative only and will vary from organization to organization. Examples of Controls and Effective Evidence can be found in the tab marked "View 3-Ex. Ctrls & Evidence" of the RCM workbook and in the “Control Objective Reference Guide.”

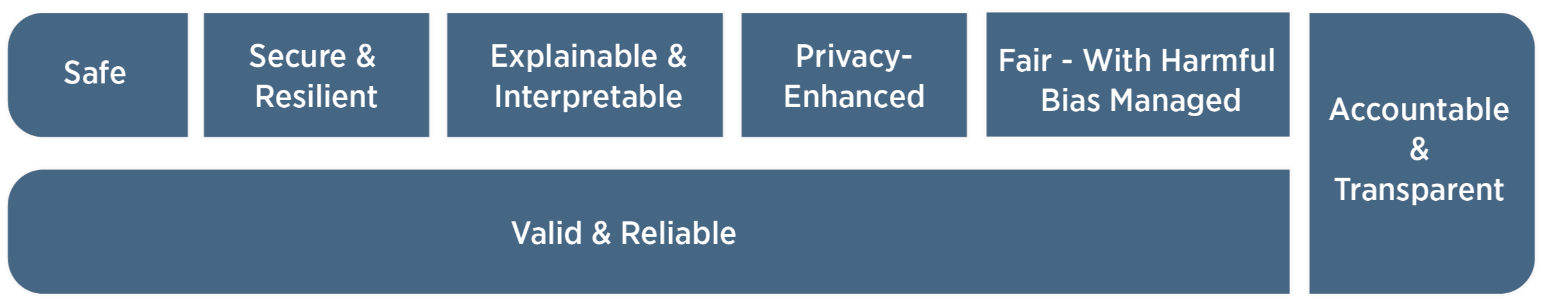
Control Objective Names and Descriptions

Below is a list of the 21 Control Objectives for Adoption Stage 1 – Initial. For additional details on each Control Objective, including its ID, Name, Description, Adoption Stage Applicability and Rationale, Implementation Guidance, sample Controls, and examples of Effective Evidence, please refer to the “RCM – Initial Stage” workbook or the “Control Objective Reference Guide.”

|  |  |  |
| --- | --- | --- |
| **Level** | **FS AI RMF Id** | **FS AI RMF Control Objective** |
| F | **GV** | **GOVERN** |
| C | **GV-1** | **Establishing Key Policies & Processes:** Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively. |
| S | **GV-1.1** | **Legal and Regulatory Compliance:** Legal and regulatory requirements involving AI are understood, managed, and documented. |
| CO | **GV-1.1.1** | **AI Legal, Regulatory, and Policy Integration:** The organization identifies, monitors, and integrates applicable laws, regulations, contractual obligations, and sector requirements into policies, procedures, and operations governing AI. This includes updating governance artifacts and aligning operational practices as requirements evolve. |
| CO | **GV-1.1.2** | **AI Compliance Responsibilities Management:** The organization clearly assigns, documents, and regularly validates its compliance responsibilities for AI-related legal, regulatory, contractual requirements, and internal organizational AI policies. |
| CO | **GV-1.1.6** | **Data Lifecycle and Retention:** The organization ensures that data retention policies, privacy requirements, and legal obligations governing data lifecycle management are integrated into its AI system lifecycle processes. It maintains records of data retention periods, manages data deletion or archiving in compliance with privacy norms, and verifies that training datasets adhere to these policies, especially when data subjects exercise their rights or when legal or privacy obligations change. |
| S | **GV-1.2** | **Defining Trustworthy AI Policies:** The characteristics of trustworthy AI are integrated into organizational policies, processes, procedures, and practices. |
| CO | **GV-1.2.1** | **AI Principles and Policy Integration:** The organization defines and documents its own AI Trustworthy Principles, which are then integrated into relevant policies, processes, and practices across key functions (e.g., HR, IT, legal, compliance). |
| CO | **GV-1.2.2** | **AI Terminology and Scope Glossary:** The organization establishes, maintains, and periodically updates a glossary of AI-related terms and concepts and clearly defines the scope, purpose, and intended/acceptable uses of its AI systems. This information is referenced and reflected consistently across organizational policies, standards, other relevant documentation, and training programs. |
| C | **GV-6** | **Establishing Third-Party Risk Management for the AI System:** Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues. |
| S | **GV-6.1** | **Establishing Third-Party Risk Management Processes:** Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third-party’s intellectual property or other rights. |
| CO | **GV-6.1.1** | **Third-Party AI Evaluation and Selection:** The organization establishes processes for evaluating and selecting third-party AI technologies based on criteria that assess security and privacy implications, integrating AI-specific considerations into third-party procurement planning, due diligence, evaluation, and contracting practices to create alignment with organizational risk management policies. |
| F | **MP** | **MAP** |
| C | **MP-1** | **Understanding the Operating Context:** Context is established and understood. |
| S | **MP-1.1** | **Defining the Operating Context:** Intended purposes, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented. Considerations include: the specific set or types of users along with their expectations; potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet; assumptions and related limitations about AI system purposes, uses, and risks across the development or product AI lifecycle; and related TEVV and system metrics. |
| CO | **MP-1.1.1** | **AI Purpose and Context Documentation:** The organization identifies and documents the AI system’s purpose, context, beneficial uses, user types and expectations, relevant laws, norms, and prospective deployment settings. |
| CO | **MP-1.1.2** | **Impact and Stakeholder Impact Assessment:** The organization considers the potential positive and negative impacts of the AI system in collaboration with a multidisciplinary set of relevant stakeholders (e.g., domain experts, human factors experts, affected communities) throughout its lifecycle. |
| CO | **MP-1.1.3** | **Assumptions, Risks, and TEVV Documentation:** The organization identifies and documents assumptions, limitations, and potential risks throughout the AI system’s lifecycle. Testing, evaluation, verification, and validation (TEVV) requirements and system metrics to identify, monitor, and manage associated system risks are identified. |
| CO | **MP-1.1.4** | **Non-AI Alternatives Evaluation:** The organization considers non-AI alternatives and documents the comparative analysis and decision rationale. |
| S | **MP-1.2** | **Bolstering Cross-Functional Perspectives on the Operating Context:** Interdisciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized. |
| CO | **MP-1.2.1** | **Multidisciplinary AI Impact Teams:** The organization recruits, supports, and facilitates documented collaboration among multidisciplinary teams throughout the AI system lifecycle (from problem formulation to monitoring). This enhances comprehensive risk analysis, bias mitigation, and thorough consideration of potential impacts. |
| S | **MP-1.3** | **Defining the Mission and Goals:** The organization’s mission and relevant goals for AI technology are understood and documented. |
| CO | **MP-1.3.1** | **AI Alignment with Mission and Values:** The organization identifies and documents the AI system’s alignment and contribution to organizational mission, values, goals (including relevant goals for AI technology), and AI Trustworthy Principles. |
| S | **MP-1.4** | **Defining Real-World Business Value:** The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated. |
| CO | **MP-1.4.1** | **AI Business Value and Context Communication:** The organization clearly defines, documents, and communicates the business value, objectives, and intended context of use for each AI system to relevant stakeholders, promoting alignment with organizational goals and facilitating transparency and accountability in how the AI system supports operational success. |
| S | **MP-1.6** | **Defining and Communicating User Requirements:** System requirements (e.g., “the system shall respect the privacy of its users”) are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks. |
| CO | **MP-1.6.1** | **AI Requirements from Stakeholders:** The organization engages relevant AI stakeholders (e.g., end-users, domain experts, impacted communities) to elicit and document system requirements that address AI risks and promote trustworthy characteristics (e.g., security, fairness, transparency, accountability). |
| C | **MP-2** | **Understanding the AI System:** Categorization of the AI system is performed |
| S | **MP-2.1** | **Defining AI System Task and Methodology:** The specific tasks and methods used to implement the tasks that the AI system will support are defined (e.g., classifiers, generative models, recommenders). |
| CO | **MP-2.1.1** | **AI Learning and Decision Tasks:** The organization defines and documents the AI system's existing and potential future learning and decision-making tasks (e.g., classification, generation, recommendation, prediction), considering fit-for-purpose, scalability and adaptability. |
| S | **MP-2.2** | **Defining AI System Utility and Oversight Needs:** Information about the AI system’s knowledge limits and how system output may be utilized and overseen by humans is documented. Documentation provides sufficient information to assist relevant AI actors when making decisions and taking subsequent actions. |
| CO | **MP-2.2.1** | **AI Usage and Boundary Documentation:** The organization documents the settings, environments, and conditions that are within the AI system’s intended use, defining the boundaries of its applicability while considering knowledge limits and areas where human intervention and irreversibility of decisions is crucial. |
| S | **MP-2.3** | **Applying Scientific Methodologies:** Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), system trustworthiness, and construct validation. |
| CO | **MP-2.3.1** | **Development and Operation Documentation:** The organization documents assumptions, limitations, techniques, and metrics used for development or operation of the AI system throughout the lifecycle, aligning with data governance policies. This includes documenting data selection, curation, preparation, and analysis techniques, as well as identifying modeled constructs and methods for inferring reasonable relationships between constructs and dataset attributes. |
| F | **MS** | **MEASURE** |
| C | **MS-1** | **Methods and Metrics:** Appropriate methods and metrics are identified and applied. |
| S | **MS-1.3** | **External Assessment of Measurement Approaches:** Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, AI actors external to the team that developed or deployed the AI system, and affected communities are consulted in support of assessments as necessary per organizational risk tolerance. |
| CO | **MS-1.3.1** | **Independent Expert Evaluation and Stakeholder Consultation:** The organization involves internal experts (external to front-line development) and independent assessors, including robust testing functions, in the regular assessment and updating of AI metrics and risk controls. Concurrently, it determines the need for consultation with domain experts, users, AI stakeholders external to the development team, and affected communities, based on risk appetite. This approach leverages multidisciplinary, unbiased insights to ensure a comprehensive understanding of potential impacts and the adequacy of risk management measures. |
| C | **MS-2** | **Evaluating AI Systems:** AI systems are evaluated for trustworthy characteristics. |
| S | **MS-2.10** | **Measuring Privacy:** Privacy risk of the AI system – as identified in the MAP function – is examined and documented. |
| CO | **MS-2.10.1** | **Initial Privacy Risk Assessment:** The organization conducts an initial examination of the privacy risks associated with its AI systems, as identified during the MAP function, and documents the results of this examination, including any identified privacy issues, concerns, or potential violations specific to AI systems. It establishes accountability mechanisms for managing AI-related privacy risks and incidents, such as data breaches or unauthorized access, and assigns clear roles and responsibilities for privacy risk management. |
| S | **MS-2.11** | **Measuring Nondiscrimination:** Fairness and bias – as identified in the MAP function – are evaluated and results are documented. |
| CO | **MS-2.11.1** | **Context-Specific Fairness and Bias Evaluation:** The organization conducts evaluations of fairness and bias in its AI systems, focusing on any regulatory requirements and the risks and potential impacts identified during the MAP function. The organization employs techniques and tools to assess fairness and bias, examining potential sources of bias throughout the AI lifecycle, including systemic bias, statistical and computational bias, and human cognitive bias. Based on the evaluation results, the organization develops and implements strategies to mitigate or manage identified biases and fairness issues. The organization documents the results of its fairness and bias evaluations, including identified issues and mitigation strategies. |
| C | **MS-3** | **Tracking AI Risks:** Mechanisms for tracking identified AI risks over time are in place. |
| S | **MS-3.3** | **Enabling Feedback, Appeal, and Override:** Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics. |
| CO | **MS-3.3.1** | **Feedback and Appeals Processes:** The organization establishes feedback processes for relevant stakeholders to report problems, concerns, or unintended consequences related to AI system outcomes, integrating with existing processes for reporting security issues and potential vulnerabilities. This includes implementing an appeals mechanism that allows users to challenge or request reviews of outcomes they believe to be inaccurate, unfair, or harmful. Both processes are documented clearly, including channels, timeframes, responsibilities, escalation procedures, and criteria for decision-making, with all reported issues tracked to closure. |

# **APPENDIX A – AI TRUSTWORTHY PRINCIPLES**

CRI leverages the NIST AI Trustworthy Principles[[6]](#footnote-7) in promoting trustworthy and responsible AI development and use. NIST's AI RMF provides guidance to organizations for managing AI risks and ensuring trustworthiness throughout the AI lifecycle. These principles are integral to building public trust and confidence in AI technologies. According to NIST, the essential characteristics of trustworthy AI, include:



1. **Validity and Reliability**: Ensuring the AI system produces accurate, consistent results and performs as intended under expected conditions. Reliability focuses on consistent performance over time, including the system's ability to learn and adapt without compromising its intended function.
2. **Safety**: AI systems should not pose risks to human life, health, property, or the environment. Designing AI for safety involves proactively mitigating potential dangers and implementing measures for safe operation, including the option to remove a system from use if necessary.
3. **Security and Resiliency**: Protecting AI systems against adversarial attacks, unauthorized access, disruptions, and ensuring they can recover from unexpected events. This includes maintaining the confidentiality, integrity, and availability of the system and its data.
4. **Accountability and Transparency**: Establishing clear responsibility for AI system outcomes and ensuring adequate information is available about the AI system's function and outputs to relevant stakeholders. Transparency involves providing clear understanding of AI processes and decisions, tailored to the user's context and needs.
5. **Explainability and Interpretability**: AI decisions should be understandable to users, and the system's reasoning process should be accessible to those affected by its output. Explainability focuses on providing reasons for an AI system's outputs, while interpretability aims to provide transparency into the AI model's architecture and how it processes data to make predictions.
6. **Privacy**: Safeguarding personal and sensitive information handled by AI systems. This involves protecting data privacy throughout the AI lifecycle, from data collection to use, sharing, and storage, and obtaining consent for data use.
7. **Fairness with Mitigation of Harmful Bias**: Ensuring equitable treatment of individuals and groups by addressing issues such as harmful algorithmic and data biases. Fairness in AI includes considering equality and equity, recognizing that mitigating bias may not guarantee overall fairness in every context.

# **APPENDIX B – ACRONYMS AND ABBREVIATIONS**

AGI Artificial General Intelligence

AI Artificial Intelligence

CIO Chief Information Officer

CISO Chief Information Security Officer

CRI Cyber Risk Institute

CMRO Chief Model Risk Officer

CRO Chief Risk Officer

ERM Enterprise Risk Management

ETL Extract Transform Load

FAQ Frequently Asked Questions

FBIIC Financial and Banking Information Infrastructure Committee

FI Financial Institution

FS Financial Services or Financial Sector

FSI Financial Services Industry

FSSCC Financial Services Sector Coordinating Counsel

GAI/GenAI Generative AI

GRC Governance, Risk, and Compliance

HITL Human-in-the-Loop

HR Human Resources

ID Identification

IP Intellectual Property

IRB Institutional/Internal Review Board

KPI Key Performance Indicator

KRI Key Risk Indicator

LLM Large Language Model

LMS Learning Management System

ML Machine Learning

MOU Memorandum of Understanding

MRM Model Risk Management

NDA Non-Disclosure Agreement

NIST National Institute of Standards and Technology

NLP Natural Language Processing

PIA Privacy Impact Assessment

PRD Product Requirements Document

PTM Pretrained Model(s)

RACI Responsible, Accountable, Consulted, and Informed [Matrix]

RAG Retrieval Augmented Generation

RIA Regulatory Impact Assessment

RCA Root Cause Analysis

RCM Risk and Control Matrix

RFI Request for Information

RFP Request for Proposal

RMF Risk Management Framework

SDLC Software Development Lifecycle

SLA Service Level Agreement

SOP Standard Operating Procedure

SQC Statistical Quality Control

TEVV Testing, Evaluation, Verification, and validation

# **APPENDIX C – GLOSSARY OF KEY TERMS**

The following key terms and their definitions are adopted from the Financial Services Sector Coordinating Council’s AI Executive Operating Group AI Lexicon and Terminology.

**Adversarial AI:** Techniques used to deceive or manipulate AI systems, causing them to make incorrect or unintended predictions or decisions. These techniques exploit vulnerabilities in AI models, often by subtly altering input data to mislead the AI system.[[7]](#footnote-8),[[8]](#footnote-9)

**Agentic AI:** A category of AI systems capable of independently making decisions, interacting with their environment, and optimizing processes without direct human intervention.[[9]](#footnote-10)

**AI Agent**: A system that autonomously perceives its environment and takes actions to achieve its goals.[[10]](#footnote-11) [[11]](#footnote-12)

**AI Model -** A component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.[[12]](#footnote-13)

**AI Drift/Decay:** The tendency for the performance of a machine learning approach to degrade over time when deployed in a real-world setting, which could result in predictions “drifting” from intended tested output and becoming less accurate and reliable as the data it encounters in production deviates from the data on which it was initially trained.[[13]](#footnote-14)

**AI Exploitation:** Techniques used to exploit vulnerabilities in AI, resulting in the compromise of data confidentiality, integrity, or availability.[[14]](#footnote-15),[[15]](#footnote-16)

**AI Risk Assessment**: The systematic process of identifying and evaluating the likelihood and potential consequences of events or actions within AI that could lead to harm.[[16]](#footnote-17)

**AI System**: An engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions that influence real or virtual environments. AI systems are designed to operate with varying levels of autonomy.[[17]](#footnote-18)

**Algorithm:** A clearly specified mathematical process for computation; a set of rules that, if followed, will give a prescribed result.[[18]](#footnote-19)

**Algorithmic Bias:** An AI system that exhibits systematically inaccurate behavior.[[19]](#footnote-20)

**Anomaly Detection System/Mechanism:** A system for identifying the occurrence of a condition that deviates from expectations based on requirements specifications, design documents, user documents, or standards, or from a person’s perceptions or experiences.[[20]](#footnote-21),[[21]](#footnote-22)

**Artificial General Intelligence:** The hypothetical intelligence of a machine that possesses the ability to understand or learn any intellectual task that a human can understand or learn; a type of AI that aims to mimic the cognitive abilities of the human brain.[[22]](#footnote-23)

**Artificial Intelligence:** A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to: (a) perceive real and virtual environments; (b) abstract such perceptions into models through analysis in an automated manner; and (c) use model inference to formulate options for information or action.[[23]](#footnote-24)

**Benchmarks**: An alternative prediction or approach used to compare a model’s inputs and outputs to estimates from alternative internal or external data or models.[[24]](#footnote-25)

**Black Box**: An AI system whose internal workings are a mystery to its users. Users can see the system’s inputs and outputs but cannot see what happens within the AI tool to produce those outputs.[[25]](#footnote-26)

**Capability Evaluation/Assessment:** A comprehensive review of an AI model’s or AI system's overall capabilities, both planned and unplanned, emerging, or malicious. Unlike specific task-focused evaluation, capability evaluation seeks to understand the full range of AI capabilities, such as how AI might adapt or evolve beyond its initial training, identifying both beneficial emergent behaviors and potential risks that could arise from autonomous operation or interaction with complex environments.[[26]](#footnote-27)

**Data Lineage**: The history of processing of a data element, which may include point-to-point data flows and the data actions performed upon the data element.[[27]](#footnote-28)

**Data Poisoning**: An attack that “corrupt[s] and contaminate[s] training data to compromise the AI system’s performance.”[[28]](#footnote-29) Such attacks can affect multiple entities simultaneously if they utilize the same poisoned dataset for training AI models.

**Data Quality/Validity**: The usefulness, accuracy, and correctness of data for its application.[[29]](#footnote-30)

**Deepfake**: AI-generated or manipulated image, audio, or video content that resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful.[[30]](#footnote-31)

**Explainability**: The extent to which AI decisioning processes and outcomes are reasonably understood by operators.[[31]](#footnote-32)

**General Purpose AI**: AI designed for use across a broad array of tasks and applications as opposed to a specific domain.[[32]](#footnote-33)

**Generative AI**: The class of AI that emulates the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.[[33]](#footnote-34)

**Guardrails:** Layered safeguards that ensure generative AI systems behave ethically, safely, and within organizational or regulatory boundaries by filtering training data, aligning model behavior, and enforcing post-deployment controls.[[34]](#footnote-35)

**Hallucinations:** A phenomenon whereby AI produces output that is erroneous or flawed but is still in the form of a convincing narrative or presentation.[[35]](#footnote-36)

**Human-in-the-Loop:** A risk-control approach for AI where a human is integrated within the AI's decision-making process. The real-time human oversight or intervention is intended to help reduce errors and strengthen accountability.[[36]](#footnote-37)

**Interpretability**: Transparency into the inner workings of AI output in the context of their designed functional purpose, helping users gain deeper insights into the functionality and trustworthiness of the system and its outputs.[[37]](#footnote-38)

**Large Language Model**: A subset of machine learning that uses algorithms trained on large amounts of data through self-supervised machine learning to recognize patterns and respond to user requests in natural language.[[38]](#footnote-39)

**Machine Learning**: An AI learning method that enables computational systems to learn patterns, make predictions, and optimize decisions from large amounts of data without being explicitly programmed for each task. Machine learning encompasses supervised, unsupervised, and reinforcement learning paradigms, serving as the technical foundation for data-driven intelligence and automation.[[39]](#footnote-40)

**Model**: A quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates. A model consists of three components: (a) information input, which delivers assumptions and data to the model; (b) processing, which transforms inputs into estimates; and (c) reporting, which translates the estimates into useful business information.[[40]](#footnote-41)

**Model Documentation**: The formal (official, approved) collection of documents and data that provides detailed explanation of the rationale, assumptions, derivations, tests, and other analyses that support the use of a model for a given purpose.[[41]](#footnote-42)

**Model Governance**: The end-to-end process by which organizations establish, implement, and maintain controls around the use of models, including everything from model documentation and version control to back-testing, model monitoring, and observability.[[42]](#footnote-43)

**Model Integrity**: The process of protecting a model against improper information modification or destruction and ensuring information non-repudiation and authenticity.[[43]](#footnote-44)

**Model Inventory**: A listing of each model, including identification and location information.[[44]](#footnote-45)

**Model Validation**: The set of processes and activities intended to verify that models are performing as expected, in line with design objectives and business uses.[[45]](#footnote-46)

**Natural Language Processing**: The analysis, interpretation, and response to human language inputs; using normal speech and writing patterns to communicate with computer systems instead of relying on programmed or pre-set responses.[[46]](#footnote-47)

**Ongoing Monitoring:** A core element of model validation that confirms a model is appropriately implemented and is performing and being used as intended.[[47]](#footnote-48)

**Output Validation**: A systematic process of verifying and confirming that AI system outputs meet specified requirements, accuracy standards, and quality criteria before being used for downstream processes.[[48]](#footnote-49)

**Performance Monitoring**: A systematic and continuous process of measuring, analyzing, and tracking AI systems’ operational effectiveness and compliance throughout their deployment lifecycle.[[49]](#footnote-50)

**Performance Threshold**: A particular value or range of values of a performance measure or diagnostic that determines the acceptance or rejection of an AI model’s performance.[[50]](#footnote-51)

**Prompt:** Natural language text describing the task that an AI should perform.[[51]](#footnote-52)

**Prompt Injection**: An attack conducted by the primary user of the system through query access.[[52]](#footnote-53)

**Responsible AI**: Conscientious design, deployment, and governance of AI systems aligned with ethical principles, societal values, and legal norms.[[53]](#footnote-54)

**Retrieval Augmented Generation:** The process of optimizing the output of a large language model, so it references an authoritative knowledge base outside its training data sources before generating a response.[[54]](#footnote-55)

**Service Level Agreement:** The specific responsibilities of a service provider [that sets and defines] the customer’s expectations.[[55]](#footnote-56)

**Structured Data:** Data that is defined and searchable, such as dates, postal codes, and merchant IDs.[[56]](#footnote-57)

**Traditional AI**: A subset of AI that focuses on performing discreet, preset tasks using predetermined algorithms and rules. These AI applications are designed to excel in a single activity or a restricted set of tasks.[[57]](#footnote-58)

**Training Data**: A subset of input data samples used to train a machine learning model.[[58]](#footnote-59)

**Unstructured Data:** Data that is more free-form, such as multimedia files, images, sound files, or unstructured text, does not necessarily follow any format or hierarchical sequence, and does not follow any relational rules.[[59]](#footnote-60)

**Validation**: Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled.[[60]](#footnote-61)

**Version Control** **(AI context):** A systematic practice of tracking, managing, and documenting changes to AI assets through their development and deployment lifecycle.[[61]](#footnote-62)

# 

# **APPENDIX D – INFORMATIVE REFERENCES**

|  |  |
| --- | --- |
| **Primary References** | |
| [NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)](https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf) | The NIST AI Risk Management Framework (AI RMF) is a voluntary framework developed by the National Institute of Standards and Technology to help organizations manage risks associated with artificial intelligence systems, promoting trustworthy and responsible AI development and deployment. |
| [NIST AI RMF Playbook](https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook) | The Playbook provides suggested actions for achieving the outcomes laid out in the AI Risk Management Framework (AI RMF) Core (Tables 1 – 4 in AI RMF 1.0). Suggestions are aligned to each sub-category within the four AI RMF functions (Govern, Map, Measure, Manage). |
| [NIST Trustworthy and Responsible AI](https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf) | A cross-sectoral profile and companion resource for the AI Risk Management Framework (AI RMF 1.0) for Generative AI, in line with Executive Order 14110 on Safe, Secure, and Trustworthy Artificial Intelligence. It aims to aid organizations in managing AI risks by integrating trustworthiness considerations into AI product and service development and offers insights and suggested actions to address novel risks associated with Generative AI. |
| [NIST AI 600-1](https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf) |
| [Supervisory Guidance on Model Risk Management (SR 11-7)](https://www.federalreserve.gov/supervisionreg/srletters/sr1107.pdf) | SR 11-7 is supervisory guidance issued by the Federal Reserve and the Office of the Comptroller of the Currency (OCC) that outlines comprehensive requirements for Model Risk Management (MRM) for banks and financial institutions operating within the United States. |
| [OECD AI Principles](https://www.oecd.org/en/topics/sub-issues/ai-principles.html) | The OECD AI Principles are the first intergovernmental standard on AI. They promote innovative, trustworthy AI that respects human rights and democratic values. Adopted in 2019 and updated in 2024, they are composed of five values-based principles and five recommendations that provide practical and flexible guidance for policymakers and AI actors. |
| [European Union (EU) AI Act](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689) | The EU Artificial Intelligence Act (AI Act) is a comprehensive legal framework established by the European Union to regulate the development and use of AI systems, aiming to address associated risks and promote trustworthy AI. |
| [CRI Profile](https://cyberriskinstitute.org/the-profile/) | The Cyber Risk Institute (CRI) Profile is a cybersecurity framework tailored for the financial sector, consolidating over 3,000 regulatory expectations into fewer than 300 diagnostic statements to streamline cyber risk management and regulatory compliance. |
| Version 2.0 of the Profile expands its scope to include enterprise technology, third-party risk management, and business continuity and resiliency, aligning with the latest NIST Cybersecurity Framework. |
| **Secondary References** | |
| [ISO 42001](file:///C:\Users\AngelaPatel\AppData\Local\Microsoft\Windows\INetCache\Documents\:b:\r\sites\US-TrustedAISolutions\Shared%20Documents\General\5.%20External%20Sources\Standards%20and%20White%20Papers\ISO%20-%20Do%20not%20copy\ISO_IEC_42001_2023(en).pdf?csf=1&web=1&e=Xb3yKE) | AI management system standard, providing organizations with a systematic approach to identify, evaluate, and address the risks associated with AI. |
| [OWASP Top 10](file:///C:\Users\AngelaPatel\AppData\Local\Microsoft\Windows\INetCache\Documents\:b:\r\sites\US-TrustedAISolutions\Shared%20Documents\General\5.%20External%20Sources\Standards%20and%20White%20Papers\OWASP\LLM_AI_Security_and_Governance_Checklist.pdf?csf=1&web=1&e=h6ndME) | The OWASP Top 10 is a standard awareness document for developers and web application security, representing a broad consensus about the most critical security risks to web applications. |
| [FFIEC Guidance](https://www.ffiec.gov/guidance/Authentication-and-Access-to-Financial-Institution-Services-and-Systems.pdf) | The Federal Financial Institutions Examination Council (FFIEC) issues guidance to standardize the supervision of financial institutions, covering areas such as cybersecurity, authentication, and information technology management. |
| [G7 Toolkit](https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/10/g7-toolkit-for-artificial-intelligence-in-the-public-sector_f93fb9fb/421c1244-en.pdf) | The G7 has established International Guiding Principles and a voluntary Code of Conduct to promote the safe, secure, and trustworthy development of advanced AI systems. |
| [Argentina](https://www.argentina.gob.ar/noticias/programa-de-transparencia-y-proteccion-de-datos-personales-en-el-uso-de-la-inteligencia) | Argentina's data protection authority, the Agency of Access to Public Information, published Resolution No. 161/23. which created the Transparency and Protection of Personal Data Program in the use of Artificial Intelligence. |
| [MAS - AI Model Risk Management](https://www.mas.gov.sg/-/media/mas-media-library/publications/monographs-or-information-paper/imd/2024/information-paper-on-ai-risk-management-final.pdf) | The Monetary Authority of Singapore (MAS) provides guidelines on the responsible use of AI and data analytics in financial services, emphasizing fairness, ethics, accountability, and transparency. |
| [UK AI Safety Institute](https://www.aisi.gov.uk/) | The United Kingdom has established the AI Safety Institute to evaluate and ensure the safety of advanced AI models, promoting responsible AI development. |
| [Germany's AI Strategy](https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf) | Germany's AI Strategy outlines the country's approach to AI, focusing on research, development, and ethical guidelines to ensure AI benefits society. |
| [Switzerland's AI Guidelines](https://www.sbfi.admin.ch/dam/sbfi/en/dokumente/2021/05/leitlinien-ki.pdf.download.pdf/leitlinien-ki_e.pdf) | Switzerland provides guidelines on the ethical use and development of AI, emphasizing transparency, accountability, and human rights. |
| [Australia's AI Ethics Framework](https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework) | Australia's framework offers principles to guide businesses and governments in the responsible development and use of AI. |
| [Japan – Social Principles of Human-Centric AI](https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf) | Japan has developed guidelines for AI governance, focusing on principles for implementing AI responsibly. |
| [Canada's Directive on Automated Decision-Making](https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592) | Canada provides a directive to ensure that automated decision systems are used responsibly, emphasizing transparency, accountability, and fairness. |
| Brazil's AI Strategy | Brazil's AI Strategy outlines the country's approach to AI, focusing on innovation, ethics, and regulatory aspects to promote AI development. |
| [Chile](https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24840) | Chile has developed a national policy for AI, aiming to promote responsible AI development and integration into various sectors. |

# **APPENDIX E – ADDITIONAL SOURCE MATERIALS**

Alsallakh, Bilal, Adeel Cheema, Chavez Procope, David Adkins, Emily McReynolds, Erin Wang, Grace Pehl, Nekesha Green, and Polina Zvyagina. 2022. Meta AI, [System-level Transparency of Machine Learning](https://ai.meta.com/research/publications/system-level-transparency-of-machine-learning/).

Bennet, Karen, Gopi Krishnan Rajbahadur, Arthit Suriyawongkul, and Kate Stewart. 2025. The Linux Foundation and SPDX, [Implementing AI Bill of Materials (AI BOM) with SPDX 3.0: A Comprehensive Guide to Creating AI and Dataset Bill of Materials.](https://arxiv.org/pdf/2504.16743)

Bryan, Pete, et al. 2025. [Taxonomy of Failure Mode in Agentic AI Systems](https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/%20Taxonomy-of-Failure-Mode-in-Agentic-AI-Systems-Whitepaper.pdf.).

Consumer Financial Protection Bureau (CFPB), May 2022. [Adverse Action Notification Requirements In Connection With Credit Decisions Based On Complex Algorithms](https://www.consumerfinance.gov/compliance/circul%20ars/circular-2022-03-%20adverse-%20action-%20notification-%20requirements-%20in-%20connection-%20with-%20credit-decisions-based-on-complex-algorithms/).

Consumer Financial Protection Bureau (CFPB), June 2023. [Chatbots in Consumer Finance](https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/).

Consumer Financial Protection Bureau (CFPB), 2024. [Official Interpretations of Regulation B (ECOA)](https://www.consumerfinance.gov/rules-policy/regulations/1002/interp-0/).

Commodity Futures Trading Commission (CFTC), May 2024. [Responsible Artificial Intelligence in Financial Markets: Opportunities, Risks & Recommendations. Commodity Futures Trading Commission, Technology Advisory Committee](https://www.cftc.gov/media/10626/TAC%20AIReport050224/download).

European Central Bank, Banking Supervision, July 2025. [ECB Guide to Internal Models. European Central Bank, Banking Supervision](https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory%20guide202507.en.%20pdf).

European Union (EU), 2024. [Artificial Intelligence Act: Annexes](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.).

Federal Housing Finance Agency (FHFA), November2013. [Model Risk Management Guidance. Advisory Bulletin, AB 2013-07](https://www.fhfa.gov/advisory-bulletin/ab-2013-07).

Federal Housing Finance Agency (FHFA), December 2022. [Advisory Bulletin 2022-03: Supplemental Guidance to AB 2013-07 – Model Risk Management Guidance](https://www.fhfa.gov/%20advisory-bulletin/AB-2022-03).

Ficklin, P.A., T. Pahl, and P. Watkins. July, 2020. CFPB Blog, [Innovation Spotlight: Providing Adverse Action Notices When Using AI/ML Models](https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models/).

Financial Industry Regulatory Authority (FINRA), June 2020. [Artificial Intelligence in the Securities Industry](https://www.finra.org/sites/default/files/2020-06/ai-report-061020.pdf).

Federal Reserve Board (FRB) Board of Governors, 2022. [Stress Tests](https://www.federalreserve.gov/supervisionreg/stress-tests-capital-planning.htm).

Federal Reserve Board (FRB) Board of Governors/Office of the Comptroller of the Currency (OCC), 2011. [Supervisory Guidance on Model Risk Management, SR 11-7; OCC 2011-12](https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf).

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wal-lach, Hal Daume III, and Kate Crawford. 2021. Communications of the ACM 64 (12): 86–92, [Datasheets for Datasets](https://dl.acm.org/doi/10.1145/3458723).

Hall, Patrick, James Curtis, and Parul Pandey. 2023. O’Reilly Media, [Machine Learning for High-risk Applications: Approaches to Responsible AI](https://www.oreilly.com/library/view/machine-learning-for/9781098102425/).

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2020. Data Protection and Privacy 12 (12): 1, [The Dataset Nutrition Label](https://arxiv.org/pdf/1805.03677).

Hubinger, Evan, 2025. SuperIntelligence-Robotics-Safety & Alignment 2 (1), [Anthropic: Responsible Scaling Policy](https://s-rsa.com/index.php/agi/article/view/13657).

International Standards Organization (ISO), 2015. [Quality management systems—Fundamentals and vocabulary](https://www.iso.org/%20standard/45481.html).

International Standards Organization (ISO), 2022. [Trustworthiness — Vocabulary](https://www.iso.org/standard/81608.html).

International Standards Organization (ISO), 2023. ISO/IEC 42001:2023, [Artificial Intelligence — Management System](https://www.iso.org/standard/81230.html).

International Standards Organization (ISO), 2025. ISO/IEC 42005:2025, [Information Technology — Artificial Intelligence (AI) — AI System Impact Assessment](https://www.iso.org/%20standard/42005).

Kraprayoon, Jam, Zoe Williams, and Rida Fayyaz. 2025. [AI Agent Governance: A Field Guide](https://arxiv.org/pdf/2505.21808).

McGregor, Sean. 2021. In Proceedings of the AAAI Conference on Artificial Intelligence, 35:15458–15463. 17, [Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database](https://arxiv.org/pdf/2011.08512).

Microsoft. 2022a. [Microsoft Responsible AI Impact Assessment Guide](https://msblogs.thesourcemediaassets.com/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Guide.pdf).

Microsoft. 2022b. [Microsoft Responsible AI Impact Assessment Template](https://msblogs.thesourcemediaassets.com/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf).

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. In Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229, [Model Cards for Model Reporting](https://dl.acm.org/doi/pdf/10.1145/3287560.3287596).

Munoz, Gary D. Lopez, Amanda J. Minnich, Roman Lutz, Richard Lundeen, Raja Sekhar Rao Dheekonda, Nina Chikanov, Bolor-Erdene Jagdagdorj, et al. 2024. PyRIT: [A Framework for Security Risk Identification and Red Teaming in Generative AI Systems](https://arxiv.org/abs/2410.02828).

National Institutes of Science and Technology (NIST), 2024. Trustworthy and Responsible AI, 600 Series,[Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile](https://nvlpubs.nist.gov/nistpubs/ai/%20NIST.AI.600-1.pdf).

Office of the Comptroller of the Currency (OCC), August 2021. [Comptroller’s Handbook: Model Risk Management.](https://www.occ.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/index-model-risk-management.html)

Office of the Comptroller of the Currency (OCC), January 2023. [Comptroller’s Handbook: Fair Lending](https://www.occ.gov/publications-and-resources/publications/comptrollers-handbo%20ok/files/fair-lending/index-fair-lending.html).

Organisation for Economic Co-operation and Development (OECD), February 2025. [Towards a Common Reporting Framework for AI Incidents](https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/towards-a-common-reporting-framework-for-ai-incidents_8c488fdb/f326d4ac-en.pdf).

Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. AI Now 9, [Algorithmic Impact Assessments: A Practical Framework for Public Agency](https://www.nist.gov/%20system/files/documents/2021/10/04/aiareport2018.pdf).

Schwartz, Reva, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, and Patrick Hall. 2022. NIST Special Publication 1270,[Towards a Standard for Identifying and Managing Bias in Artificial Intelligence](https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/towards-a-common-reporting-framework-for-ai-incidents_8c488fdb/f326d4ac-en.pdf).

Sudjianto, Agus, Aijun Zhang, Srinivas Neppalli, Tarun Joshi, and Michal Malohlava. 2024. [Human-Calibrated Automated Testing and Validation of Generative Language Models](https://arxiv.org/pdf/2411.16391).

1. The CRI Profile and Cloud Profile can be found on the CRI website: <https://cyberriskinstitute.org/the-profile/> [↑](#footnote-ref-2)
2. The FS AI RMF can be found and downloaded at no cost on the CRI website: [www.cyberriskinstitute.org](http://www.cyberriskinstitute.org). [↑](#footnote-ref-3)
3. No part of the FS AI RMF is considered itself a risk assessment or the basis of a risk assessment. [↑](#footnote-ref-4)
4. For a complete list of primary and secondary sources, please see Appendix D. [↑](#footnote-ref-5)
5. U.S. Department of the Treasury Report on the Uses, Opportunities, and Risks of Artificial Intelligence in Financial Services: <https://home.treasury.gov/news/press-releases/jy2760> [↑](#footnote-ref-6)
6. For more information on the NIST AI Trustworthy Principles please see: https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/#:~:text=Characteristics%20of%20trustworthy%20AI%20systems,fair%20with%20harmful%20bias%20managed. [↑](#footnote-ref-7)
7. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf> [↑](#footnote-ref-8)
8. <https://www.nccoe.nist.gov/ai/adversarial-machine-learning> [↑](#footnote-ref-9)
9. <https://doi.org/10.1016/j.array.2025.100399> [↑](#footnote-ref-10)
10. <https://www.ibm.com/think/topics/ai-agents> [↑](#footnote-ref-11)
11. [ISO/IEC DIS 22989(en), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology](https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:dis:ed-1:v1:en) [↑](#footnote-ref-12)
12. <https://csrc.nist.gov/glossary/term/artificial_intelligence_model#:~:text=Definitions:,HEADQUARTERS> [↑](#footnote-ref-13)
13. <https://www.ibm.com/think/topics/model-drift> [↑](#footnote-ref-14)
14. <http://www.ibm.com/think/topics/model-drift> [↑](#footnote-ref-15)
15. <https://www.belfercenter.org/publication/AttackingAI> [↑](#footnote-ref-16)
16. [Terms and definitions | Voluntary AI Safety Standard | Department of Industry Science and Resources](https://www.industry.gov.au/publications/voluntary-ai-safety-standard/terms-and-definitions) [↑](#footnote-ref-17)
17. [Glossary - AIRC](https://airc.nist.gov/glossary/) [↑](#footnote-ref-18)
18. <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-107r1.pdf> [↑](#footnote-ref-19)
19. [nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf](https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf) [↑](#footnote-ref-20)
20. [NIST SP 800-160v1r1](https://doi.org/10.6028/NIST.SP.800-160v1r1) from [ISO/IEC/IEEE 24765:2017](https://www.iso.org/standard/71952.html) [↑](#footnote-ref-21)
21. [NIST SP 800-37 Rev. 2](https://doi.org/10.6028/NIST.SP.800-37r2) under detect (CSF function) from [NIST Cybersecurity Framework Version 1.1](https://doi.org/10.6028/NIST.CSWP.04162018) [↑](#footnote-ref-22)
22. <https://cloud.google.com/discover/what-is-artificial-general-intelligence> [↑](#footnote-ref-23)
23. [15 USC 9401: Definitions](https://uscode.house.gov/view.xhtml?req=(title:15%20section:9401%20edition:prelim)) [↑](#footnote-ref-24)
24. [Model Risk Management, Comptroller's Handbook](https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf) [↑](#footnote-ref-25)
25. [What Is Black Box AI and How Does It Work? | IBM](https://www.ibm.com/think/topics/black-box-ai) [↑](#footnote-ref-26)
26. [Terms and definitions | Voluntary AI Safety Standard | Department of Industry Science and Resources](https://www.industry.gov.au/publications/voluntary-ai-safety-standard/terms-and-definitions) [↑](#footnote-ref-27)
27. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf [↑](#footnote-ref-28)
28. <https://www.bis.org/fsi/publ/insights63.pdf> [↑](#footnote-ref-29)
29. [NIST Big Data Interoperability Framework: Volume 4, Security and Privacy](https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-4r2.pdf) [↑](#footnote-ref-30)
30. [Glossary - AIRC](https://airc.nist.gov/glossary/) [↑](#footnote-ref-31)
31. Adapted from, [Model Risk Management, Comptroller's Handbook](https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf), p. 103 [↑](#footnote-ref-32)
32. Adapted from, <https://www.bis.org/fsi/publ/insights63.pdf> [↑](#footnote-ref-33)
33. [USCODE-2023-title15-chap119-sec9401.pdf](https://www.govinfo.gov/content/pkg/USCODE-2023-title15/pdf/USCODE-2023-title15-chap119-sec9401.pdf) [↑](#footnote-ref-34)
34. [What Are AI Guardrails? Ensuring Safe and Ethical Generative AI - Mindgard](https://mindgard.ai/blog/what-are-ai-guardrails) [↑](#footnote-ref-35)
35. Adapted from <https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf> and <https://home.treasury.gov/system/files/261/FSOC2024AnnualReport.pdf> [↑](#footnote-ref-36)
36. Adapted from, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> [↑](#footnote-ref-37)
37. Adapted from, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> [↑](#footnote-ref-38)
38. <https://home.treasury.gov/system/files/261/FSOC2024AnnualReport.pdf> [↑](#footnote-ref-39)
39. Adapted from, <https://www.fsb.org/uploads/P011117.pdf> and <https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf> [↑](#footnote-ref-40)
40. <https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf> [↑](#footnote-ref-41)
41. <https://www.openriskmanual.org/wiki/Model_Documentation> [↑](#footnote-ref-42)
42. [What Is Model Governance? | IBM](https://www.ibm.com/think/topics/model-governance) [↑](#footnote-ref-43)
43. <https://www.nccoe.nist.gov/publication/1800-25/VolA/index.html> [↑](#footnote-ref-44)
44. [Small Business Information Security: the Fundamentals](https://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.7621r1.pdf) [↑](#footnote-ref-45)
45. [Model Risk Management, Comptroller's Handbook](https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf) [↑](#footnote-ref-46)
46. [Summary Report: Standards Requirements Gathering Workshop for Natural Language Analysis](https://nvlpubs.nist.gov/nistpubs/ams/NIST.AMS.100-30.pdf) [↑](#footnote-ref-47)
47. [Model Risk Management, Comptroller's Handbook](https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf) [↑](#footnote-ref-48)
48. Adapted from, [AI Test, Evaluation, Validation and Verification (TEVV) | NIST](https://www.nist.gov/ai-test-evaluation-validation-and-verification-tevv) [↑](#footnote-ref-49)
49. [Artificial Intelligence Risk Management Framework (AI RMF 1.0)](https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf) [↑](#footnote-ref-50)
50. [Model Risk Management, Comptroller's Handbook](https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf) [↑](#footnote-ref-51)
51. [Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector](https://home.treasury.gov/system/files/136/Managing-Artificial-Intelligence-Specific-Cybersecurity-Risks-In-The-Financial-Services-Sector.pdf) [↑](#footnote-ref-52)
52. Adapted from, [direct prompting attack - Glossary | CSRC](https://csrc.nist.gov/glossary/term/direct_prompting_attack) [↑](#footnote-ref-53)
53. [Artificial Intelligence Risk Management Framework (AI RMF 1.0)](https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf) [↑](#footnote-ref-54)
54. [What is RAG? - Retrieval-Augmented Generation AI Explained - AWS](https://aws.amazon.com/what-is/retrieval-augmented-generation/) [↑](#footnote-ref-55)
55. [CNSSI 4009](https://www.cnss.gov/CNSS/openDoc.cfm?a=ZzQpuWVxCe9cbduemCczmQ%3D%3D&b=182777ABC0ECF2B8C958324742AB098B04DD76BC73C110B953C7497B6724E9FAE561754EF8C1E26BF2F46CF099C94507), p.184 [↑](#footnote-ref-56)
56. [6250\_PA\_Glossary\_AI\_Terms](https://thepaymentsassociation.org/wp-content/uploads/2024/06/6250_PA_Glossary_AI_Terms_V5-FINAL.pdf) [↑](#footnote-ref-57)
57. <https://www.uschamber.com/co/run/technology/traditional-ai-vs-generative-ai> [↑](#footnote-ref-58)
58. [ISO/IEC DIS 22989(en), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology](https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:dis:ed-1:v1:en) [↑](#footnote-ref-59)
59. [Glossary: Unstructured Data | resources.data.gov](https://resources.data.gov/glossary/unstructured-data/) [↑](#footnote-ref-60)
60. <https://www.iso.org/standard/45481.html> [↑](#footnote-ref-61)
61. Adapted from, [Artificial Intelligence Risk Management Framework (AI RMF 1.0)](https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf) [↑](#footnote-ref-62)