US Government Releases Artificial Intelligence Governance Framework

Client memorandum | January 31, 2023

Authors: Amir R. Ghavi, Katelyn E. James, and Cecily D'Amore

On January 26, 2023, the US government, through the National Institute of Standards and Technology ("NIST")[1], released an artificial intelligence ("AI") governance framework, titled the Artificial Intelligence Risk Management Framework ("AI RMF") 1.0. The AI RMF is the result of a multiyear process involving workshops and industry participation and feedback, with the goal of mitigating risks in the design, development, use and evaluation of AI products, services and systems. The AI RMF was designed with two primary goals: (i) to help increase trustworthiness of AI and (ii) to manage risks associated with the development and use of AI. NIST intends to finalize its draft guidebook to the AI RMF, called the AI RMF Playbook, in the Spring of 2023.[2]

The AI RMF 1.0 is divided into two parts: (I) Foundational Information and (II) Core and Profiles. Part I addresses how organizations should consider framing risks related to their AI systems, including:

- Understanding and addressing the risk, impact and harm that may be associated with AI systems.
- Addressing the challenges for AI risk management, including those related to third-party software, hardware and data.
- Incorporating a broad set of perspectives across the AI life cycle.

Part I also describes trustworthy AI systems, including characteristics such as validity and reliability, safety, security and resilience, accountability and transparency, explainability and interpretability, privacy-enhanced, and fairness with harmful bias managed.

Part II describes features to address risks associated with the use and deployment of AI systems. These features include:

- Governance: a culture of risk management;
- Mapping: context is recognized and risks identified;
- Measurement: identified risks are assessed, analyzed or tracked; and
- Management: risks are prioritized and acted upon based on a projected impact.

The AI RMF is not legally binding or required for AI development and deployment, but will likely become a de facto standard for AI governance.[3] We have seen in 2022 the awesome power of AI to extend human capability, creativity and insights. But in the absence of governance, AI systems[4] (like all disruptive technologies) pose potential risks. Researchers have highlighted latent risks in raw data sets used to train the AI systems and unintended consequences in the use and operation of AI systems.[5] The AI RMF is intended to address risks and equip AI actors[6] to manage such risks in a responsible way to enhance trustworthiness and ultimately cultivate public trust in the AI systems. As AI continues to evolve, NIST intends for the AI RMF to evolve with it to reflect new knowledge, awareness and practices. In this memorandum, we will summarize the AI RMF for organizations who currently use or plan to use AI in the future.

I. Foundational Information: AI Risks and Trustworthiness

The AI RMF states that trustworthy systems must be responsive to a variety of criteria. As trustworthiness is inextricably connected to social and organizational behavior, the AI RMF recommends that humans guide the specific metrics related to AI trustworthiness. However, the AI RMF acknowledges that a comprehensive approach to risk management must recognize tradeoffs. The AI RMF takes the position that a trustworthy AI system is: (a) valid and reliable, (b) safe, (c) secure and resilient, (d) accountable and transparent, (e) explainable and interpretable, (f) privacy-enhanced, and (g) fair with harmful bias managed.

Fig. 4. Characteristics of trustworthy AI systems. Valid & Reliable is a necessary condition of trustworthiness and is shown as the base for other trustworthiness characteristics. Accountable & Transparent is shown as a vertical box because it relates to all other characteristics.

Source*: NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0)*

(a) *Valid and Reliable*: The AI RMF purports that the measurement of validity, accuracy, robustness and reliability (detailed below) contribute to trustworthiness. It suggests ongoing testing or monitoring to confirm that a system performs as intended, prioritizing minimization of potential negative impacts, and employing human intervention as necessary where the AI system cannot detect or correct errors.

- *Validation*: The confirmation that requirements for an intended use or application have been fulfilled through objective evidence can decrease negative AI risks and increase trustworthiness.

- *Accuracy:* Measures of accuracy should consider false positive and false negative rates and human-AI teaming, and demonstrate validity that goes beyond the training conditions. They should be clearly defined and documented, and include details about test methodology that are representative of conditions of expected use.
- *Robustness or generalizability:* Being able to maintain an AI system's level of performance under a variety of circumstances, including those not initially anticipated, can minimize potential harms.
- *Reliability:* A goal of overall correctness of AI system operation under the conditions of expected use and over a given period of time, including the lifetime of the system, can contribute to an AI system's trustworthiness.

(b) *Safe*: The AI RMF encourages safe operation of AI systems through tailored AI risk management based on the context and severity of potential risks presented. It states that those approaches should start early in the AI system's life cycle and should allow for the ability to shut down, modify or have human intervention incorporated into systems that deviate from intended or expected functionality. Additionally, the AI RMF contends that safe AI systems are improved through responsible design, development and deployment; clear information to deploys on responsible uses of the system; responsible decision-making by deployers and end users; and explanations and documentation of risk based on empirical evidence of incidents.

(c) *Secure and Resilient*: The AI RMF considers security and resilience to be related but distinct characteristics. It suggests that resilience requires the maintenance of normal functionality in the face of adverse or unexpected events or changes in their environment, while security includes the protocols to avoid, protect against, respond to or recover from attacks. It contends that AI systems may be secure if they can maintain confidentiality, integrity, and availability through protection mechanisms that prevent unauthorized access and use.

(d) *Accountable and Transparent:* The AI RMF takes the position that trustworthy AI depends on accountability, and accountability presupposes transparency. It states that transparency should involve tailoring how access to information is provided based on the stage of the AI life cycle and the role or knowledge of AI actors or those interacting with or using the AI system, and that this information may include design decisions, training data, model structure, the model's intended use cases, and how and when deployment, post-deployment, or end-use decisions were made and by whom. The AI RMF recommends developers test different types of transparency tools to ensure that AI systems are used as intended.

(e) *Explainable and Interpretable:* In the AI RMF, explainabilty refers to a representation of the mechanisms underlying AI systems' operation, whereas interpretability refers to the meaning of AI systems' output in the context of their designed functional purposes. The AI RMF suggests that this information can help end users understand the purposes and potential impact of an AI system, and that risks should be managed by tailoring descriptions to individual differences and by communicating why the AI system made certain predictions or recommendations.

(f) *Privacy-Enhanced:* The AI RMF takes the position that privacy helps safeguard human autonomy, identity and dignity, and that AI system decisions should be guided by privacy values such as anonymity, confidentiality and control. The AI RMF states that privacy-related risks may influence security, bias and transparency, and that AI systems may introduce new risks to privacy, including inferences that may identify individuals or information that was previously private.

(g) *Fair – with Harmful Bias Managed:* The AI RMF reflects the perspective that bias goes beyond data representativeness and demographic balance, and is tightly associated with fairness in society. Although perceptions of fairness differ and may shift depending on application, the AI RMF contends that fairness in AI is rooted in concerns for equality and equity. It suggests that organizations' risk management efforts should recognize and consider the differences that may impact AI systems. The AI RMF identifies several categories of AI bias to be considered and managed: systemic, computational and statistical, and human-cognitive.

- *Systemic Bias*. Systemic bias can impact AI systems at various levels, ranging from an AI dataset to the broader society that uses AI systems.
- *Computational and Statistical Bias*. This form of bias may be present in AI datasets and algorithmic processes, stemming from systematic errors from non-representative samples.
- *Human-Cognitive Bias*. How individuals perceive AI system information and make decisions, as well as how they think about purposes and functions of an AI system, may impact many stages of the AI life cycle, including its design, implementation, operation and maintenance.

II. Core and Profiles

AI RMF Core

The AI RMF Core suggests the outcomes and actions that are meant to enable dialogue, understanding and activities necessary to manage AI risks. The Core is comprised of three elements: (i) functions, (ii) categories and (iii) subcategories. The four functions organize AI risk management activities at their highest level to govern, map, measure and manage AI risks. The categories and subcategories subdivide the function into specific outcomes and actions. The AI RMF Core functions should be implemented to reflect diverse and multidisciplinary perspectives, and may be applied differently among different organizations to manage risk based on resources and capabilities. However, we note that while NIST advocates that the AI RMF Core should include views from outside of the organization, this may be impractical for organizations trying to maintain trade secrets or the confidentiality of their products and services.

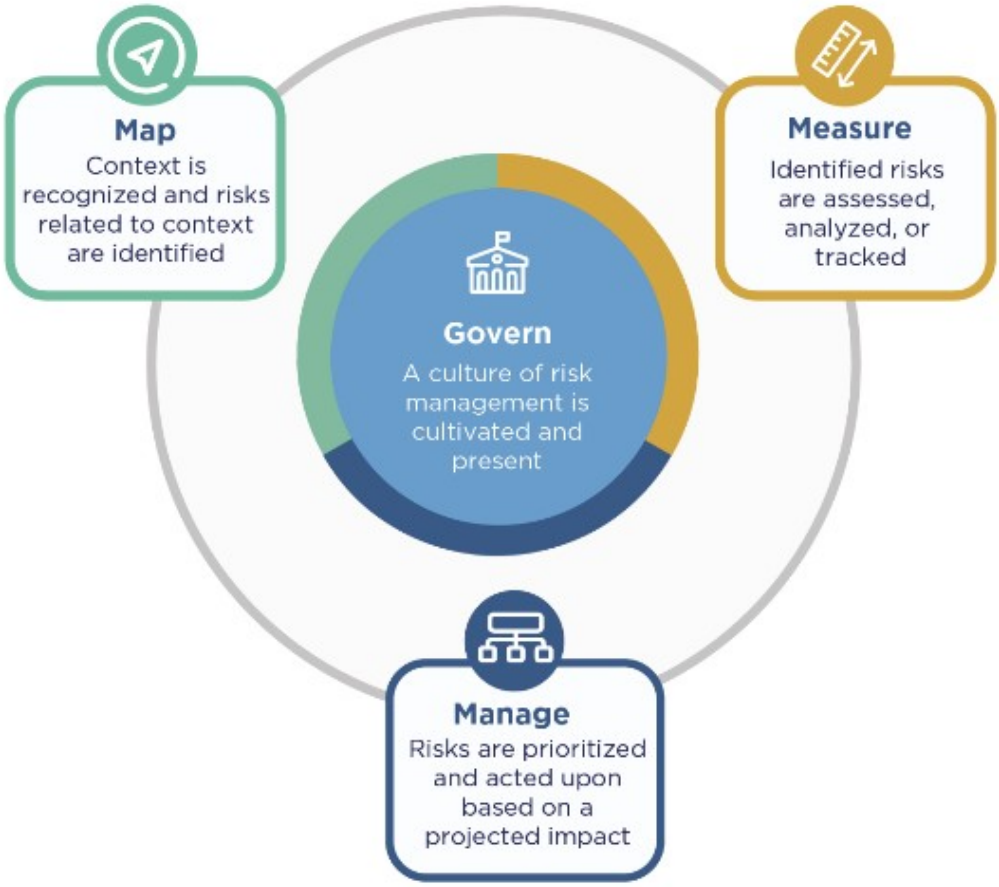

Fig. 5. Functions organize AI risk management activities at their highest level to govern, map, measure, and manage AI risks. Governance is designed to be a cross-cutting function to inform and be infused throughout the other three functions.

Source*: NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0)*

Govern: The Govern function is designed to cultivate a culture of risk management within organizations. The AI RMF establishes that governance focuses on both technical aspects of AI system design and development and on organizational practices and competencies that directly affect the individuals involved in training, deploying and monitoring such systems.

The AI RMF suggests the Govern function is cross-cutting, and enables the other functions of the AI risk management process, especially influencing those related to compliance or evaluation. It takes the position that governance is a continual and intrinsic requirement for effective AI risk management over an AI system's lifespan. The AI RMF states that governance provides a structure through which AI risk management functions can better align with organizational policies and strategic priorities, including those that do not directly relate to AI systems. Practices related to governing AI risks are described in the NIST AI RMF Playbook.

The March 2022 workshop hosted by NIST highlighted the potential harms of poorly governed AI development and deployment in high-stakes areas such as banking, transportation, criminal justice and employment.

NIST has provided example categories and subcategories to incorporate the Govern function:

Category	Subcategory
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.	• Legal and regulatory requirements involving AI are understood, managed and documented. • The characteristics of trustworthy AI are integrated into organizational policies, processes, procedures and practices. • Processes, procedures and practices are in place to determine the needed level of risk management activities based on the organization's risk tolerance. • The risk management process and its outcomes are established through transparent policies, procedures and other controls based on organizational risk priorities. • Ongoing monitoring and periodic review of the risk management process and its outcomes are planned and organizational roles and responsibilities clearly defined, including determining the frequency of periodic review. • Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities. • Processes and procedures are in place for decommissioning and phasing out AI systems safely and in a manner that does not increase or decrease the organization's trustworthiness.

Category	Subcategory
Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible and trained for mapping, measuring and managing AI risks.	• Roles and responsibilities and lines of communication related to mapping, measuring and managing AI risks are documented and are clear to individuals and teams throughout the organization. • The organization's personnel and partners receive AI risk management training to enable them to perform their duties and responsibilities consistent with related policies, procedures and agreements. • Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.
Workforce diversity, equity, inclusion and accessibility processes are prioritized in the mapping, measuring and managing of AI risks throughout the life cycle.	• Decision-making related to mapping, measuring and managing AI risks throughout the life cycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise and backgrounds). • Policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems.
Organizational teams are committed to a culture that considers and communicates AI risk.	• Organizational policies and practices are in place to foster a critical thinking and safety-first mindset in the design, development, deployment and uses of AI systems to minimize negative impacts. • Organizational teams document the risks and potential impacts of the AI technology they design, develop, deploy, evaluate and use, and they communicate about the impacts more broadly. • Organizational practices are in place to enable AI testing, identification of incidents and information sharing.

Category	Subcategory
Processes are in place for robust engagement with relevant AI actors.	• Organizational policies and practices are in place to collect, consider, prioritize and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks. • Mechanisms are established to enable the team that developed or deployed AI systems to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation.
Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues.	• 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. • Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.

Map: The Map function establishes context to frame risks related to an AI system. Outcomes in the Map function inform both the Measure and Manage functions. The AI RMF acknowledges that the diversity of actors and activities at various stages of an AI system's life cycle can make it difficult to anticipate impacts of AI systems, and it is likely that AI actors in charge of one part of an AI system may not have full visibility or control into another part of the AI system. Further, the AI RMF suggests that information gathered while carrying out this function can inform decisions about model management, including an initial decision about whether an AI solution is necessary. The AI RMF encourages incorporating both perspectives from a diverse internal team, and also engagement with those external to the team that developed or deployed the AI system, and articulates that such perspectives are critical in implementing this function, as they help organizations proactively prevent risks and, in turn, develop more trustworthy AI systems.

The AI RMF recommends that the implementation of the Map function incorporate perspectives from internal and external teams that have developed or deployed the AI system, as well as external collaborators, end users and potentially impacted communications. The AI RMF states that such perspectives may help organizations prevent negative risk and develop more trustworthy AI systems by improving their ability to understand context, identify the limitations of AI processes, and anticipate risk of the use of AI beyond the intended use.

NIST provided examples of the categories and the subcategories for the Mapping function:

Category	Subcategory
Context is established and understood.	• 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 use 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 life cycle; and related test, evaluation, verification and validation ("TEVV") and system metrics. • 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. • The organization's mission and relevant goals for the AI technology are understood and documented. • The business value or context of business use has been clearly defined or, in the case of assessing existing AI systems, re-evaluated. • Organizational risk tolerances are determined and documented. • System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks.
Classification of the AI system is performed.	• The specific tasks and methods used to implement the tasks that the AI system will support are defined (e.g., classifiers, generative models, recommenders). • 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.

Category	Subcategory
	• 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.
AI capabilities, targeted usage, goals, and expected benefits and costs compared with appropriate benchmarks are understood.	• Potential benefits of intended AI system functionality and performance are examined and documented. • Potential costs, including non-monetary costs, which result from expected or realized AI errors or system functionality and trustworthiness, as connected to organizational risk tolerance, are examined and documented. • Targeted application scope is specified and documented based on the system's capability, established context and AI system categorization. • Processes for operator and practitioner proficiency with AI system performance and trustworthiness, and relevant technical standards and certifications, are defined, assessed and documented. • Processes for human oversight are defined, assessed and documented in accordance with organizational policies from the Govern function.
Risks and benefits are mapped for all components of the AI system, including third-party software and data.	• Approaches for mapping AI technology and the legal risks of its components, including the use of third-party data or software, are in place, followed and documented, as are the risks of infringement of a third party's intellectual property or other rights. • Internal risk controls for components of the AI system, including third-party AI technologies, are identified and documented.

Category	Subcategory
Impacts to individuals, groups, communities, organizations and society are characterized.	• Likelihood and magnitude of each identified impact (both potentially beneficial and harmful), based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system or other data, are identified and documented. • Practices and personnel for supporting regular engagement with relevant AI actors and integrating feedback about positive, negative and unanticipated impacts are in place and documented.

Measure: The Measure function helps organizations determine the knowledge relevant to AI risks, including tracking metrics for the aforementioned trustworthy characteristics, social impact and human-AI configurations. Under the AI RMF, the measure function includes quantitative, qualitative, or mixed-method assessment and analysis to monitor AI risks and their impacts. The AI RMF suggests that the methodologies should adhere to scientific, legal and ethical norms, and are to be carried out in an open and transparent process. The AI RMF establishes that measuring AI risks includes tracking metrics for the trustworthy characteristics of AI systems, the social impact of such systems and human-AI configurations. It states that measurement can provide a traceable basis to inform management decisions when tradeoffs among the trustworthy characteristics arise. Upon completion of this function, the AI RMF takes the position that objective, repeatable or scalable TEVV processes will be in place, followed and documented. Practices related to measuring AI risks are described in the NIST AI RMF Playbook.

NIST has provided sample categories and subcategories for the Measure function:

Category	Subcategory
Appropriate methods and metrics are identified and applied.	• Approaches and metrics for measurement of AI risks enumerated during the Map function are selected for implementation starting with the most significant AI risks. The risks or trustworthiness characteristics that will not, or cannot, be measured are properly documented. • Appropriateness of AI metrics and effectiveness of existing controls are regularly assessed and updated, including reports of errors and potential impacts on affected communities.

Category	Subcategory
	• Internal experts who did not serve as frontline 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.
AI systems are evaluated for trustworthy characteristics.	• Test sets, metrics and details about the tools used during TEVV are documented. • Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population. • AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment settings. Measures are documented. • The functionality and behavior of the AI system and its components, as identified in the Map function, are monitored when in production. • The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented. • The AI system is evaluated regularly for safety risks, as identified in the Map function. The AI system to be deployed is demonstrated to be safe, its residual negative risk does not exceed risk tolerance, and it can fail safely, particularly if made to operate beyond its knowledge limits. Safety metrics reflect system reliability and robustness, real-time monitoring and response times for AI system failures. • AI system security and resilience, as identified in the Map function, are examined and documented. • Risks associated with transparency and accountability, as identified in the Map function, are examined and documented.

Category	Subcategory
	• The AI model is explained, validated and documented, and AI system output is interpreted within its context, as identified in the Map function, to inform responsible use and governance. • Privacy risk of the AI system, as identified in the Map function, is examined and documented. • Fairness and bias, as identified in the Map function, are evaluated and results are documented. • Environmental impact and sustainability of AI model training and management activities, as well as the privacy risk of the AI system, as identified in the Map function, are assessed and documented. • Effectiveness of the employed TEVV metrics and processes in the Measure function are evaluated and documented.
Mechanisms for tracking identified AI risks over time are in place.	• Approaches, personnel and documentation are in place to regularly identify and track existing, unanticipated and emergent AI risks, based on factors such as intended and actual performance in deployed contexts. • Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available. • Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.
Feedback about efficacy of measurement is gathered and assessed.	• Measurement approaches for identifying AI risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.

Category	Subcategory
	• Measurement results regarding AI system trustworthiness in deployment context(s) and across the AI life cycle are informed by input from domain experts and relevant AI actors to validate whether the system is performing consistently as intended. Results are documented. • Measurable performance improvements or declines based on consultations with relevant AI actors, including affected communities, and field data about context-relevant risks and trustworthiness characteristics are identified and documented.

Manage: The Manage function ties together all four functions by allocating risk management resources on a regular basis as defined by the Govern function. According to the AI RMF, it addresses the risks that have been mapped and measured in order to maximize the benefits of AI systems and minimize any adverse impacts. It states that contextual information previously gathered and already-established systemic documentation practices are also utilized in this function to bolster risk management efforts. The AI RMF urges Framework users to continue to apply the Manage function to deployed AI systems as methods, contexts, risks, needs and expectations all evolve over time. Practices related to managing AI risks are described in the NIST AI RMF Playbook.

NIST has provided sample categories and subcategories for prioritizing the Manage function:

Category	Subcategory
AI risks based on assessments and other analytical output from the Map and Measure functions are prioritized, responded to and managed.	• A determination is made as to whether the AI system achieves its intended purpose and stated objectives and whether its development or deployment should proceed. • Treatment of documented AI risks is prioritized based on impact, likelihood and available resources or methods. • Responses to the AI risks deemed high priority, as identified by the Map function, are developed, planned and documented. Risk response options can include mitigating, transferring, avoiding or accepting.

Category	Subcategory
Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented and informed by input from relevant AI actors.	• Resources required to manage AI risks are taken into account, along with viable non-AI alternative systems, approaches or methods, to reduce the magnitude or likelihood of potential impacts. • Mechanisms are in place and applied to sustain the value of deployed AI systems. • Procedures are followed to respond to and recover from a previously unknown risk when it is identified. • Mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.
AI risks and benefits from third-party entities are managed.	• AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented. • Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.
Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.	• Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response and change management. • Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI actors.

AI RMF Profiles

The AI RMF has established domain-tailored profiles for the purpose of implementing the AI RMF functions. The AI RMF use-case profiles, such as a *hiring profile* or *fair housing profile*, are implementations of AI RMF functions, categories and subcategories for a specific setting or application. These profiles may illustrate how risk can be managed at a certain stage of the AI life cycle or in a specific sector, technology or end-use application. AI RMF temporal profiles describe either the current or the desired target state of specific AI risk management activities in a given sector, industry, organization or application context. An AI RMF current profile indicates how AI is currently being managed and the related risks in terms of current outcomes, whereas a target profile indicates the outcomes needed to achieve the desired or target AI risk management goals, and comparing the two can reveal gaps to address. AI RMF cross-sectoral profiles cover risks of models or applications that can be used across use cases or sectors. To achieve greater flexibility, the AI RMF does not prescribe profile templates.

---

[1] NIST, a part of the Department of Commerce, promotes US innovation and industrial competitiveness by advancing measurement science, standards and technology in ways that enhance economic security and improve quality of life.

[2] The AI RMF Playbook suggests ways to use and develop the AI RMF 1.0. The playbook is still in draft form and NIST is seeking public comments until February 27, 2023. Feedback can be sent to AIframework@nist.gov and the draft AI RMF Playbook can be accessed here.

[3] In 2014, NIST released the Cybersecurity Framework, a voluntary framework that establishes comprehensive cybersecurity and information security practices. Absent federal standard or regulations, the Cybersecurity Framework has become the de facto standard for commercially reasonable cybersecurity practices. Absent federal standard or regulations, the Cybersecurity Framework has become the de facto standard for what is considered commercially reasonable cybersecurity practices. The Cybersecurity Framework has since been adopted by federal agencies and governments, as well as private entities and organizations. The Cybersecurity Framework is also recognized internationally and is available in ten languages.

[4] The 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 that are designed to operate with varying levels of autonomy.

[5] For example, the Federal Trade Commission ("FTC") was tasked by Congress to complete a study on how AI can be used to address online harms. In their June 2022 report to Congress, *Combatting Online Harms Through Innovation,* the FTC addresses how datasets that support AI tools are often "not robust or accurate enough to avoid false positives or false negatives." Additionally, in 2020, the FTC released guidance on its blog regarding the commercial use of AI and algorithms, titled "Using Artificial Intelligence and Algorithms." The guidance outlined several recommendations for businesses to take when deploying AI into the market organized around four key values: (i) transparency, (ii) fairness, (iii) accuracy and (iv) accountability. In 2021, the FTC offered additional insight regarding the use of AI, in a second piece of guidance, titled "Aiming for truth, fairness, and equity in your company's use of AI," which focused on examples when AI practices can be deceptive or unfair, building upon its recommendations from 2020.

[6] The AI RMF refers to "AI actors" as individuals and organizations who deploy or operate AI
