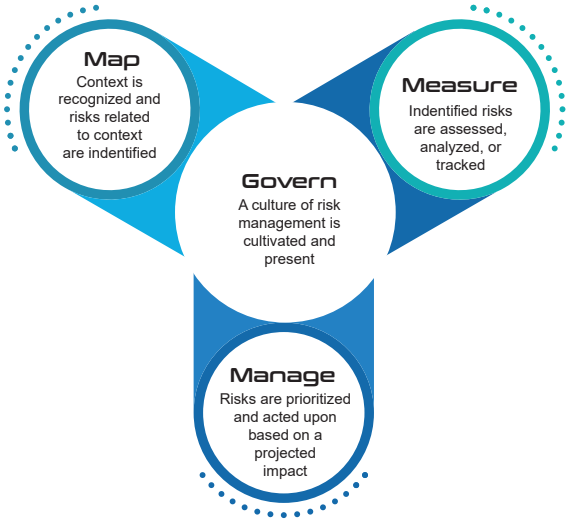
An Infographic

AI and NIST Cybersecurity

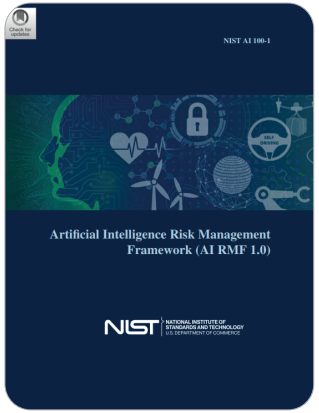
AI Risk Management Framework (RMF)

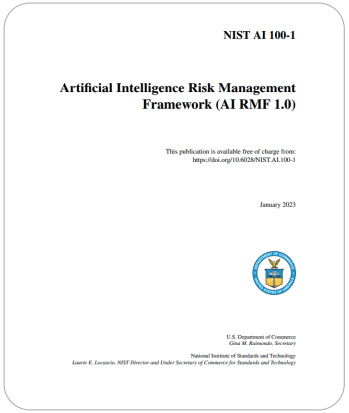
- AI technologies can drive inclusive economic growth and support scientific advancements.
- AI risk management can drive responsible uses and practices by prompting organizations and their internal teams who design, develop, and deploy AI to think more critically about context and potential or unexpected negative and positive impacts.
- AI RMF is intended to help developers, users and evaluators of AI systems.
- The Framework users and AI actors should consider and encompass trustworthiness characteristics during pre-design, design and development, deployment, use, and test and evaluation of AI technologies and systems.

AI RMF Core

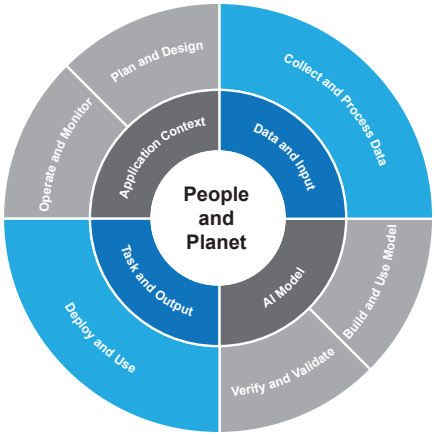


Source





Lifecycle and Key Dimensions of an AI System



Potential Harms Related to AI Systems

Harm to People

- Individual: Harm to a person's civil liberties, rights, physical or psychological safety, or economic opportunity.
- Group/Community: Harm to a group such as discrimination against a population sub-group.
- Societal: Harm to democratic participation or educational access.

Harm to an Organization

- Harm to an organizations business operations.
- Harm to an organization from security breaches or monetary loss.
- Harm to an organization's reputation.

Harm to an Ecosystem

- Harm to interconnected and interdependent elements and resources.
- Harm to the global financial system, supply chain, or interrelated systems.
- Harm to natural resources, the environment, and planet.

Challenges for AI Risk Management

ecfirst

AI Defense, *Beyond Cyber*

Global Cyber Defense Thought Leader

© ecfirst. All Rights Reserved. 2024

An Infographic

AI and NIST Cybersecurity

AI Functions

GOVERN

- GOVERN is a cross-cutting function that is infused throughout AI risk management and enables the other functions of the process.
- Attention to governance is a continual and intrinsic requirement for effective AI risk management over an AI system's lifespan and the organization's hierarchy.
- Strong governance can drive and enhance internal practices and norms to facilitate organizational risk culture.
- Management aligns the technical aspects of AI risk management to policies and operations.
- Senior leadership sets the tone for risk management within an organization.
- Documentation can enhance transparency, improve human review processes, and bolster accountability in AI system teams.

Map

- The MAP function establishes the context to frame risks related to an AI system. The AI lifecycle consists of many interdependent activities involving a diverse set of actors.
- The information gathered while carrying out the MAP function enables negative risk prevention and informs decisions for processes such as model management, as well as an initial decision about appropriateness or the need for an AI solution.
- Outcomes in the MAP function are the basis for the MEASURE and MANAGE functions. Without contextual knowledge, and awareness of risks within the identified contexts, risk management is difficult to perform.
- Implementation of this function is enhanced by incorporating perspectives from a diverse internal team and engagement with those external to the team that developed or deployed the AI system.
- Gathering such broad perspectives can help organizations proactively prevent negative risks and develop more trustworthy AI systems by:
  - Improving their capacity for understanding contexts
  - Checking their assumptions about context of use
  - Enabling recognition of when systems are not functional within or out of their intended context.
  - Identifying positive and beneficial uses of their existing AI systems.
  - Improving understanding of limitations in AI and Machine Learning (ML) processes.
  - Identifying constraints in real-world applications that may lead to negative impacts.
  - Identifying known and foreseeable negative impacts related to intended use of AI systems.
  - Anticipating risks of the use of AI systems beyond intended use.
- After completing the MAP function, Framework users should have sufficient contextual knowledge about AI system impacts to inform an initial go/no-go decision about whether to design, develop, or deploy an AI system.

Measure

- The Measure function employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts.
- Processes developed or adopted in the MEASURE function should include rigorous software testing and performance assessment methodologies with associated measures of uncertainty, comparisons to performance benchmarks, and formalized reporting and documentation of results.
- Measurement provides a traceable basis to inform management decisions.
- Options may include recalibration, impact mitigation, or removal of the system from design, development, production, or use, as well as a range of compensating, detective, deterrent, directive, and recovery controls.
- After completing the MEASURE function, objective, repeatable, or scalable TEVV processes including metrics, methods, and methodologies are in place, followed, and documented.

Manage

- The MANAGE function entails allocating risk resources to mapped and measured risks on a regular basis and as defined by the GOVERN function.
- Risk treatment comprises plans to respond to, recover from, and communicate about incidents or events.
- Contextual information gleaned from expert consultation and input from relevant AI actors – established in GOVERN and carried out in MAP – is utilized in this function to decrease the likelihood of system failures and negative impacts.
- Systematic documentation practices bolster AI risk management efforts and increase transparency and accountability.
- After completing the MANAGE function, plans for prioritizing risk and regular monitoring and improvement will be in place.
- Framework users will have enhanced capacity to manage the risks of deployed AI systems and to allocate risk management resources based on assessed and prioritized risks.
- It is incumbent on Framework users to continue to apply the MANAGE function to deployed AI systems as methods, contexts, risks, and needs or expectations from relevant AI actors evolve over time.

© ecfirst. All Rights Reserved. 2024

An Infographic

AI and NIST Cybersecurity

AI Function 1: Govern

Categories	Subcategories
GOVERN 1 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.	GOVERN 1.1 Legal and regulatory requirements involving AI are understood, managed, and documented. GOVERN 1.2 The characteristics of trustworthy AI are integrated into organizational policies, processes, procedures, and practices. GOVERN 1.3 Processes, procedures, and practices are in place to determine the needed level of risk management activities based on the organization's risk tolerance. GOVERN 1.4 The risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities. GOVERN 1.5 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. GOVERN 1.6 Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities. GOVERN 1.7 Processes and procedures are in place for decommissioning and phasing out AI systems safely and in a manner that does not increase risks or decrease the organization's trustworthiness.
GOVERN 2 Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.	GOVERN 2.1 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. GOVERN 2.2 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. GOVERN 2.3 Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.

Categories	Subcategories
GOVERN 3 Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle.	GOVERN 3.1 Decision-making related to mapping, measuring, and managing AI risks throughout the lifecycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds). GOVERN 3.2 Policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems.
GOVERN 4 Organizational teams are committed to a culture that considers and communicates AI risk.	GOVERN 4.1 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 potential negative impacts. GOVERN 4.2 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. GOVERN 4.3 Organizational practices are in place to enable AI testing, identification of incidents, and information sharing.
GOVERN 5 Processes are in place for robust engagement with relevant AI actors.	GOVERN 5.1 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. GOVERN 5.2 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.
GOVERN 6 Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues.	GOVERN 6.1 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. GOVERN 6.2 Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.

© ecfirst. All Rights Reserved. 2024

An Infographic

AI and NIST Cybersecurity

AI Function 2: Map

Categories	Subcategories
MAP 1 Context is established and understood.	MAP 1.1 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 Test, Evaluation, Verification, and Validation (TEVV) and system metrics.
	MAP 1.2 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.
	MAP 1.3 The organization's mission and relevant goals for AI technology are understood and documented.
	MAP 1.4 The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated.
	MAP 1.5 Organizational risk tolerances are determined and documented.
	MAP 1.6 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.
MAP 2 Context is established and understood.	MAP 2.1 The specific tasks and methods used to implement the tasks that the AI system will support are defined (e.g., classifiers, generative models, recommenders).
	MAP 2.2 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.
	MAP 2.3 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.

Categories	Subcategories
MAP 3 AI capabilities, targeted usage, goals, and expected benefits and costs compared with appropriate benchmarks are understood.	MAP 3.1 Potential benefits of intended AI system functionality and performance are examined and documented.
	MAP 3.2 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.
	MAP 3.3 Targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization.
	MAP 3.4 Processes for operator and practitioner proficiency with AI system performance and trustworthiness – and relevant technical standards and certifications – are defined, assessed, and documented.
	MAP 3.5 Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from the GOVERN function.
MAP 4 Risks and benefits are mapped for all components of the AI system including third-party software and data.	MAP 4.1 Approaches for mapping AI technology and legal risks of its components – including the use of third-party data or software – are in place, followed, and documented, as are risks of infringement of a third party's intellectual property or other rights.
	MAP 4.2 Internal risk controls for components of the AI system, including third-party AI technologies, are identified and documented.
MAP 5 Impacts to individuals, groups, communities, organizations, and society are characterized.	MAP 5.1 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.
	MAP 5.2 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.

© ecfirst. All Rights Reserved. 2024

An Infographic

AI and NIST Cybersecurity

AI Function 3: Measure

Categories	Subcategories	
Measure 2 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.	Measure 1.1	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.
	Measure 1.2	Appropriateness of AI metrics and effectiveness of existing controls are regularly assessed and updated, including reports of errors and potential impacts on affected communities.
	Measure 1.3	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.
Measure 2 AI systems are evaluated for trustworthy characteristics.	Measure 2.1	Test sets, metrics, and details about the tools used during TEVV are documented.
	Measure 2.2	Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population.
	Measure 2.3	AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s). Measures are documented.
	Measure 2.4	The functionality and behavior of the AI system and its components – as identified in the MAP function – are monitored when in production.
	Measure 2.5	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.
	Measure 2.6	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 the 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.
	Measure 2.7	AI system security and resilience – as identified in the MAP function – are evaluated and documented.
	Measure 2.8	Risks associated with transparency and accountability – as identified in the MAP function – are examined and documented.

Categories	Subcategories	
Measure 2 AI systems are evaluated for trustworthy characteristics.	Measure 2.9	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.
	Measure 2.10	Privacy risk of the AI system – as identified in the MAP function – is examined and documented.
	Measure 2.11	Fairness and bias – as identified in the MAP function – are evaluated and results are documented.
	Measure 2.12	Environmental impact and sustainability of AI model training and management activities – as identified in the MAP function – are assessed and documented.
	Measure 2.13	Effectiveness of the employed TEVV metrics and processes in the MEASURE function are evaluated and documented.
Measure 3 Mechanisms for tracking identified AI risks over time are in place.	Measure 3.1	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.
	Measure 3.2	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.
	Measure 3.3	Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.
Measure 4 Feedback about efficacy of measurement is gathered and assessed.	Measure 4.1	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.
	Measure 4.2	Measurement results regarding AI system trustworthiness in deployment context(s) and across the AI lifecycle are informed by input from domain experts and relevant AI actors to validate whether the system is performing consistently as intended. Results are documented.
	Measure 4.3	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.

© ecfirst. All Rights Reserved. 2024

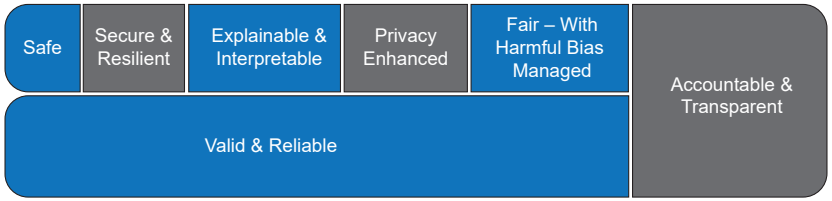
An Infographic

AI and NIST Cybersecurity

AI Function 4: Manage

Categories	Subcategories	
Manage 1 AI risks based on assessments and other analytical output from the MAP and MEASURE functions are prioritized, responded to, and managed.	Manage 1.1	A determination is made as to whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed.
	Manage 1.2	Treatment of documented AI risks is prioritized based on impact, likelihood, and available resources or methods.
	Manage 1.3	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.
	Manage 1.4	Negative residual risks (defined as the sum of all unmitigated risks) to both downstream acquirers of AI systems and end users are documented.
Manage 2 Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors.	Manage 2.1	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.
	Manage 2.2	Mechanisms are in place and applied to sustain the value of deployed AI systems.
	Manage 2.3	Procedures are followed to respond to and recover from a previously unknown risk when it is identified.
	Manage 2.4	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.
Manage 3 AI risks and benefits from third-party entities are managed.	Manage 3.1	AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.
	Manage 3.2	Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.
Manage 4 Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.	Manage 4.1	Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.
	Manage 4.2	Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI actors.
	Manage 4.3	Incidents and errors are communicated to relevant AI actors, including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.

Characteristics of Trustworthy AI Systems



AI RMF Resources

ecfirst

AI Defense, *Beyond Cyber*

Global Cyber Defense Thought Leader

© ecfirst. All Rights Reserved. 2024