NIST AI 100.1
NIST AI Risk Management Framework
(AI RMF 1.0)

NIST AI 100-1
NIST AI Risk Management Framework

- Published January 2023
- Offers a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems.
- Is intended to be:
  - ***Voluntary***, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the Framework
  - Practical, to adapt to the AI landscape as AI technologies continue to develop, and to be operationalized by organizations in varying degrees and capacities so society can benefit from AI while also being protected from its potential harms
  - Flexible and to augment existing risk practices which should align with applicable laws, regulations, and norms
- Is designed to equip organizations and individuals – referred to here as *AI actors* – with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, deployment, and use of AI systems over time
- Offers approaches to minimize anticipated negative impacts of AI systems *and* identify opportunities to maximize positive impacts
- Designed to address new risks as they emerge

NIST AI 100-1
NIST AI Risk Management Framework
Attributes of the Framework

- Be risk-based, resource-efficient, pro-innovation, and voluntary
- Be Consensus-driven and developed and regularly updated through an open, transparent process
- Uses clear and plain language that is understandable by a broad audience, including senior executives, government officials, non-governmental organization leadership, and those who are not AI professionals – while still of sufficient technical depth to be useful to practitioners
- Provide common language and understanding to manage AI risks
- Be easily usable and fit well with other aspects of risk management
- Be useful to a wide range of perspectives, sectors, and technology domains
- Be outcome-focused and non-prescriptive
- Take advantage of and foster greater awareness of existing standards, guidelines, best practices, methodologies, and tools for managing AI risks – as well as illustrate the need for additional, improved resources
- Be law- and regulation-agnostic
- Be a living document

NIST AI 100-1
NIST AI Risk Management Framework
Some Key Definitions

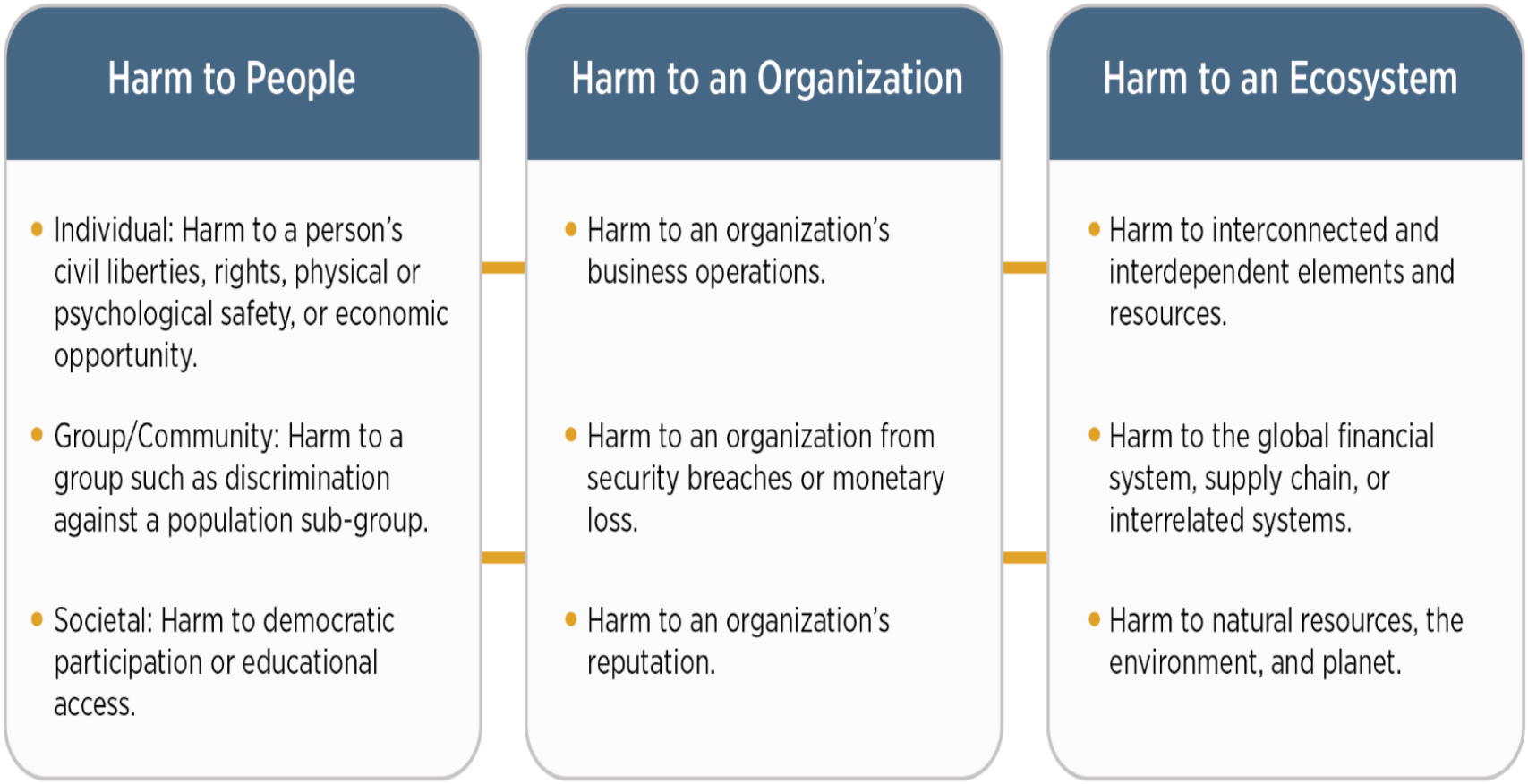
- **AI actors**: Those who play an active role in the AI system lifecycle, including organizations and individuals that deploy or operate AI (OECD (2019) Artificial Intelligence in Society—OECD iLibrary)
- **Artificial Intelligence (AI) System**: An engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022)
- **Risk**: The composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event. The impacts, or consequences, of AI systems can be positive, negative, or both and can result in opportunities or threats (Adapted from: ISO 31000:2018)
- **Risk Management**: Coordinated activities to direct and control an organization with regard to risk (Source: ISO 31000:2018)
- **Risk Tolerance**: The organization's or stakeholder's readiness or appetite to bear the risk in order to achieve its objectives
- **Accuracy**: *C*loseness of results of observations, computations, or estimates to the true values or the values accepted as being true (ISO/IEC TS 5723:2022)

*OECD - Organisation for Economic Co-operation and Development

NIST AI 100-1
NIST AI Risk Management Framework
Some Key Definitions

- **Robustness or generalizability**: Ability of an AI system to maintain its level of performance under a variety of circumstances
- **Social responsibility**: The organization's responsibility "for the impacts of its decisions and activities on society and the environment through transparent and ethical behavior" (ISO 26000:2010)
- **Sustainability**: The "state of the global system, including environmental, social, and economic aspects, in which the needs of the present are met without compromising the ability of future generations to meet their own needs" (ISO/IEC TR 24368:2022)
- "**Professional Responsibility**": An approach that "aims to ensure that professionals who design, develop, or deploy AI systems and applications or AI-based products or systems, recognize their unique position to exert influence on people, society, and the future of AI" (ISO/IEC TR 24368:2022)
- **TEVV**: Test, Evaluation, Verification and Validation
- **Validation**: The "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (Source: ISO 9000:2015)
- **Reliability**: The "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (Source: ISO/IEC TS 5723:2022)

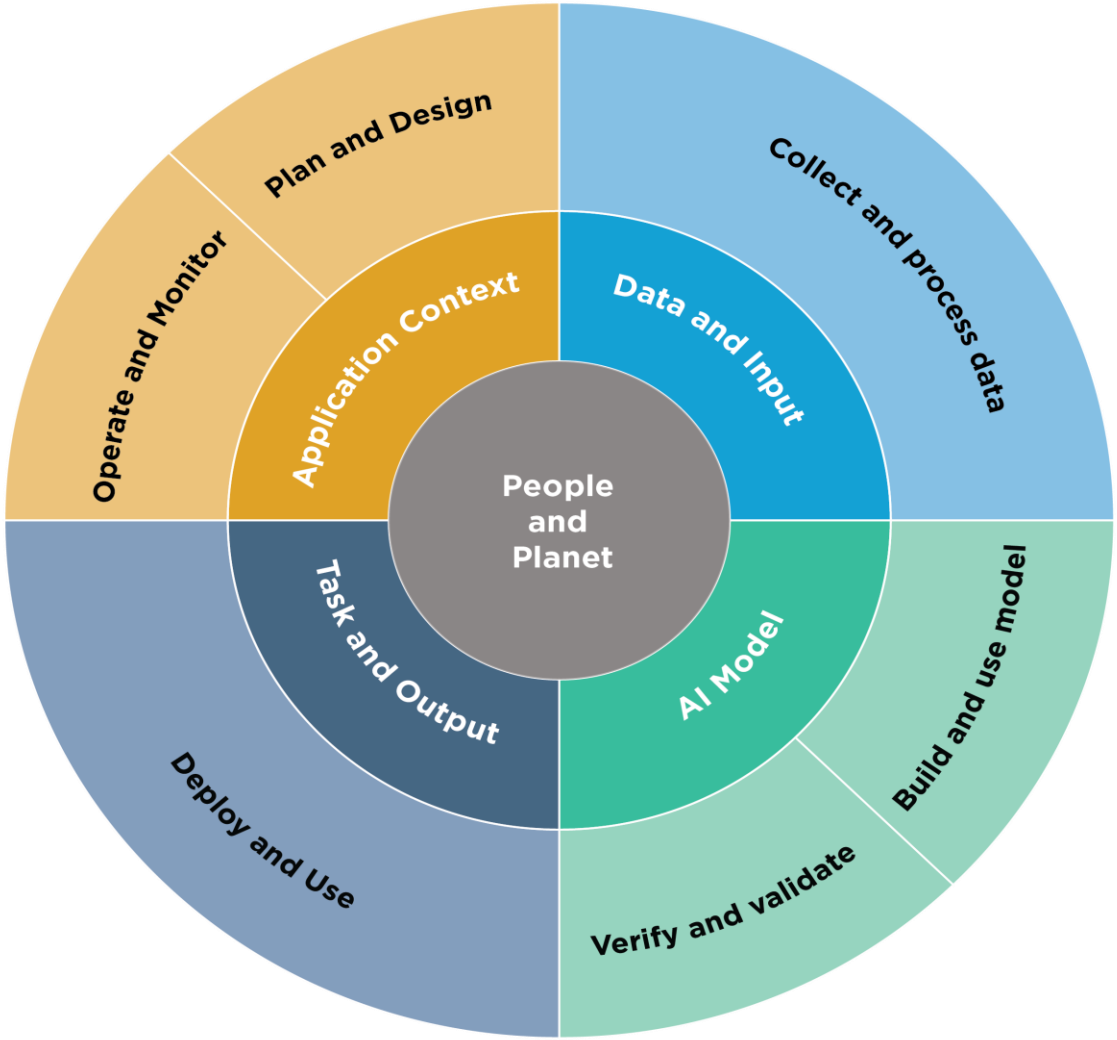
NIST AI 100-1
NIST AI Risk Management Framework
Examples of Potential Harms Related to AI Systems



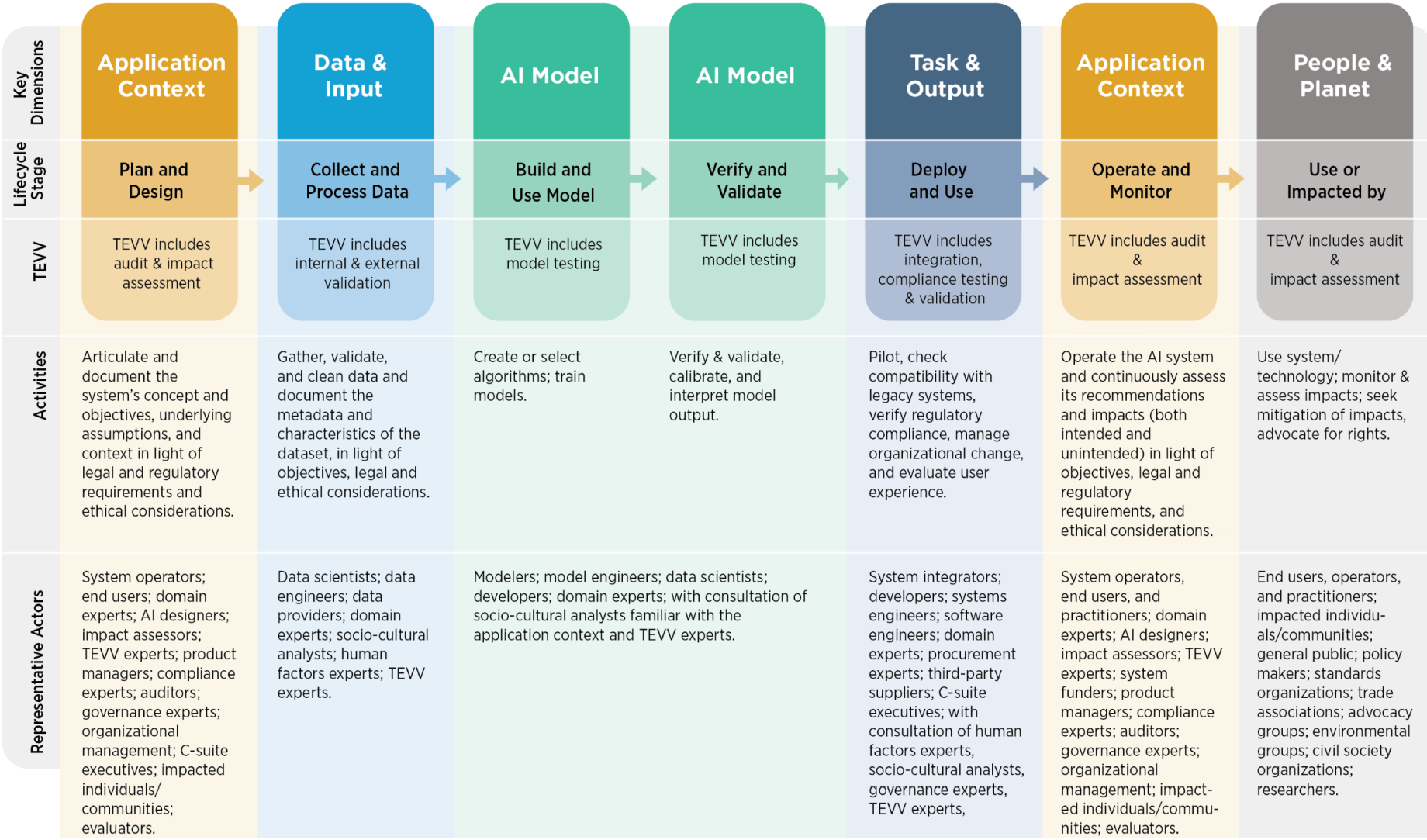
NIST AI Risk Management Framework
Risk Management Challenges

- AI risks and impacts that are not well-defined or adequately understood are difficult to measure quantitatively or qualitatively
- Cannot prescribe risk tolerance – need to equip organizations to define reasonable risk tolerance, manage those risks, and document their risk management process
- Attempting to eliminate risk entirely can be counterproductive in practice – because incidents and failures cannot be eliminated – and may lead to unrealistic expectations and resource allocation that may exacerbate risk and make risk triage impractical
- AI risks should not be considered in isolation; Need to integrate AI risks with other critical risks

NIST AI 100-1
NIST AI Risk Management Framework
Lifecycle and Key Dimensions of an AI System



NIST AI 100-1
NIST AI Risk Management Framework
AI Actors Across AI Lifecycle Stages



NIST AI 100-1
NIST AI Risk Management Framework
Characteristics of Trustworthiness

Trustworthy AI is: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and *Fair with Harmful Bias Managed*

- Approaches which enhance AI trustworthiness can also contribute to a reduction of AI risks
- Addressing AI trustworthy characteristics individually will not assure AI system trustworthiness, and tradeoffs are always involved
- Increasing the breadth and diversity of stakeholder input throughout the AI lifecycle can enhance opportunities for identifying AI system benefits and positive impacts, and increase the likelihood that risks arising in social contexts are managed appropriately

NIST AI 100-1
NIST AI Risk Management Framework Characteristics of Trustworthiness

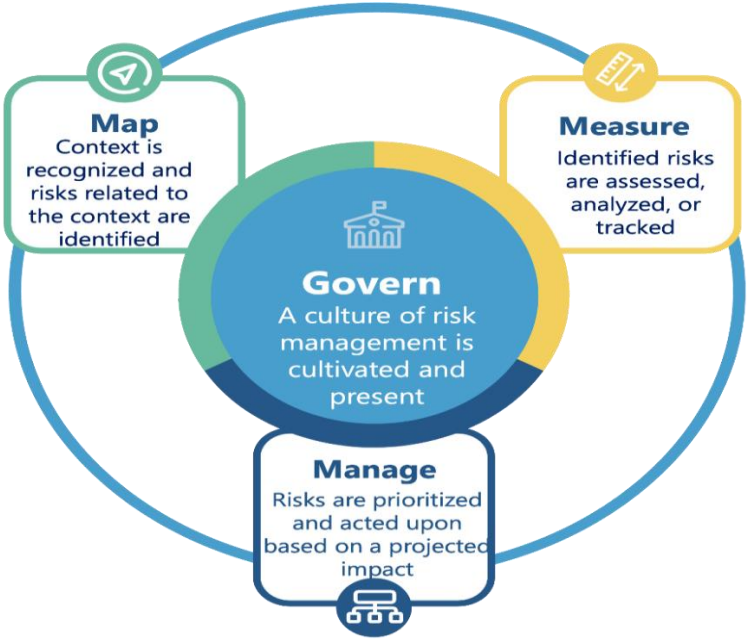
- **Valid and Reliable –** Means that the system is performing as intended without failure for a specified time period under a variety of circumstances
- **Safe** - AI systems should not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered
- **Secure and Resilient** - AI systems must be able to (1) withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary and (2) maintain confidentiality, integrity, and availability
- **Transparent and Accountable** – Reflects the extent to which information is available to individuals about an AI system, if they are interacting – or even aware that they are interacting – with such a system
- **Explainable and Interpretable** – Explainability refers to providing a representation of the mechanisms underlying an algorithm's operation, whereas interpretability refers to the meaning of AI systems' output in the context of its designed functional purpose
- **Privacy-Enhanced** - Norms and practices that help to safeguard human autonomy, identity, and dignity
- **Fair – *With Harmful Bias Managed*** - Includes concerns for equality and equity by addressing issues such as bias and discrimination

NIST AI 100-1
NIST AI Risk Management Framework
Benefit to Users

- Enhanced processes for governing, mapping, measuring, and managing AI risk, and clearly documenting outcomes;
- Improved awareness of the relationships and tradeoffs among trustworthiness characteristics, socio-technical approaches, and AI risks;
- Explicit processes for making go/no-go system commissioning and deployment decisions;
- Established policies, processes, practices, and procedures for improving organizational accountability efforts related to AI system risks;
- Enhanced organizational culture which prioritizes the identification and management of AI system risks and potential impacts to individuals, communities, organizations, and society;
- Better information sharing within and across organizations about risks, decision-making processes, responsibilities, common pitfalls, TEVV practices, and approaches for continuous improvement;
- Greater contextual knowledge for increased awareness of downstream risks;
- Strengthened engagement with interested parties and relevant AI actors; and
- Augmented capacity for TEVV of AI systems and associated risks

NIST AI 100-1
NIST AI Risk Management Framework
Core Functions

- **Govern**: Cultivate and implement a culture of risk management within organizations developing, deploying, or acquiring AI systems
- **Map**: Establish the context to frame risks related to an AI system
- **Measure**: Employ quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts
- **Manage**: Entails allocating risk management resources to mapped and measured risks on a regular basis and as defined by the Govern function



NIST AI 100-1
NIST AI Risk Management Framework
Core Steps – Govern Function

- Cultivates and implements a culture of risk management within organizations designing, developing, deploying, evaluating, or acquiring AI systems;
- Outlines processes, documents, and organizational schemes that anticipate, identify, and manage the risks a system can pose, including to users and others across society – and procedures to achieve those outcomes;
- Incorporates processes to assess potential impacts;
- Provides a structure by which AI risk management functions can align with organizational principles, policies, and strategic priorities;
- Connects technical aspects of AI system design and development to organizational values and principles, and enables organizational practices and competencies for the individuals involved in acquiring, training, deploying, and monitoring such systems; and
- Addresses full product lifecycle and associated processes, including legal and other issues concerning use of third-party software or hardware systems and data

NIST AI 100-1
NIST AI Risk Management Framework
Govern Function 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, and procedures.
	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 ~~1.3~~ 1.4: The risk management process and its outcomes are established through transparent ~~mechanisms and all significant risks as determined are measured~~ *policies procedures, and other controls based on organizational risk priorities*.
	GOVERN ~~1.4~~ 1.5: Ongoing monitoring and periodic review of the risk management process and its outcomes are planned~~, with~~ *and* organizational roles and responsibilities *are* 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.

NIST AI 100-1
NIST AI Risk Management Framework
Govern Function Categories/Subcategories

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.*

NIST AI 100-1
NIST AI Risk Management Framework
Govern Function 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 ~~demographically and disciplinarily~~ diverse team ~~including internal and external personnel~~ *(e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds)*. ~~Specifically, teams that are directly engaged with identifying design considerations and risks include a diversity of experience, expertise, and backgrounds to ensure AI systems meet requirements beyond a narrow subset of users.~~
	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.

NIST AI 100-1
NIST AI Risk Management Framework
Govern Function Categories/Subcategories

GOVERN 4: Organizational teams are committed to a culture that considers and communicates risk.	GOVERN 4.1: Organizational *policies and* practices are in place to foster a critical thinking and safety-first mindset in the design, development, ~~and~~ 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, ~~or~~ 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.

NIST AI 100-1
NIST AI Risk Management Framework
Govern Function Categories/Subcategories

GOVERN 5: Processes are in place for robust stakeholder engagement.	GOVERN 5.1: Organizational policies and practices are in place to collect, consider, prioritize, and integrate *feedback from those* external ~~stakeholder feedback~~ *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 ~~AI actors~~ *the team that developed or deployed AI systems* to regularly incorporate adjudicated ~~stakeholder~~ feedback *from relevant AI actors* into system design and implementation.
GOVERN 6: Policies and procedures are in place to address AI risks 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.

NIST AI 100-1
NIST AI Risk Management Framework
Core Steps – Map Function

Helps 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 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; and
- Anticipating risks of the use of AI systems beyond intended use

NIST AI 100-1
NIST AI Risk Management Framework
Map Function Categories/Subcategories

MAP 1: Context is established and understood.	MAP 1.1: Intended purpose, prospective settings in which the AI system will be deployed, the specific set or types of users along with their expectations, and impacts of system use 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*.
	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~~ 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 ~~1.4~~ 1.3: The organization's mission and relevant goals for the AI technology are understood.

NIST AI 100-1
NIST AI Risk Management Framework
Map Function Categories/Subcategories

MAP 1: Context is established and understood.	MAP 1.5: Organizational risk tolerances are determined.
	~~MAP 1.6: Practices and personnel for design activities enable regular engagement with stakeholders, and integrate actionable user and community feedback about unanticipated negative impacts.~~
	MAP ~~1.7~~ 1.6: System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood from stakeholders. Design decisions take socio-technical implications into account to address AI risks.

NIST AI 100-1
NIST AI Risk Management Framework
Map Function Categories/Subcategories

MAP 2: ~~Classification~~ *Categorization* of the AI system is performed.	MAP 2.1: The specific task, and methods used to implement the task, that the AI system will support is 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), and construct validation.
MAP 3: AI capabilities, targeted usage, goals, and expected benefits and costs compared with the status quo are understood.	MAP 3.1: 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 performance *and trustworthiness – as connected to organizational risk tolerance* – are examined and documented.
	MAP 3.3: Targeted application scope is specified~~, narrowed,~~ and documented based on *the system's capability,* established context and AI system ~~classification~~ *categorization*.

NIST AI 100-1
NIST AI Risk Management Framework
Map Function Categories/Subcategories

MAP 3: AI capabilities, targeted usage, goals, and expected benefits and costs compared with the status quo are understood.	*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 ~~third-party~~ ~~technology~~ *AI and legal* risks *of its components – including the use of third-party data or software –* are in place 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* risks are ~~in place~~ *identified* and documented.

NIST AI 100-1
NIST AI Risk Management Framework
Map Function Categories/Subcategories

MAP 5: Impacts to individuals, groups, communities, organizations, and society are ~~assessed~~ *characterized*.	~~MAP 5.1: Potential positive and negative impacts to individuals, groups, communities, organizations, and society are regularly identified and documented.~~
	MAP ~~5.2~~ 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, ~~stakeholder~~ feedback *from those external to the team that developed or deployed the AI system*, or other data are identified and documented.
	~~MAP 5.3: Assessments of benefits versus impacts are based on analyses of impact, magnitude, and likelihood of risk.~~
	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.*

NIST AI 100-1
NIST AI Risk Management Framework
Core Steps – Measure Function

- Employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts
- Uses knowledge relevant to AI risks identified in the **MAP** function and informs the **MANAGE** function
- Includes tracking metrics for trustworthy characteristics, social impact, and human-AI configurations
- 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

NIST AI 100-1
NIST AI Risk Management Framework
Measure Function Categories/Subcategories

MEASURE 1: Appropriate methods and metrics are identified and applied.	MEASURE 1.1: Approaches and metrics for ~~quantitative or qualitative~~ measurement of *AI* risks~~,~~ ~~identified by the outcome of~~ *enumerated during* the Map function~~,~~ *are selected for implementation starting with the most significant AI risks* ~~including context-relevant measures of trustworthiness are identified and selected for implementation~~. 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 is 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 external stakeholders~~ and affected communities are consulted in support of assessments *as necessary per organizational risk tolerance*.

NIST AI 100-1
NIST AI Risk Management Framework
Measure Function Categories/Subcategories

Category	Subcategory
MEASURE 2: 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 ~~comply with~~ *meet applicable* requirements *(including* human subject protection*)*~~;~~ and ~~human subjects or datasets~~ are representative of the ~~intended~~ 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 ~~2.4~~ 2.5: Deployed product is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.
	MEASURE ~~2.5~~ 2.6: AI system is evaluated regularly for safety. *The AI system to be* deployed ~~product~~ is demonstrated to be safe*, its residual negative risk does not exceed the risk tolerance,* and *it* can fail safely, particularly if ~~and gracefully if it is~~ made to operate beyond its knowledge limits. Safety metrics reflect ~~implicate~~ system reliability and robustness, real-time monitoring, and response times for AI system failures.
	~~MEASURE 2.6: Computational bias is evaluated regularly and results are documented.~~
	MEASURE 2.7: AI system resilience and security – *as identified in the* ***MAP*** *function* – is evaluated regularly and documented.

NIST AI 100-1
NIST AI Risk Management Framework
Measure Function Categories/Subcategories

Category	Subcategory
MEASURE 2: Systems are evaluated for trustworthy characteristics.	MEASURE 2.8: *Risks associated with transparency and accountability – as identified in the **MAP** function – are examined and documented.*
	MEASURE ~~2.8~~ 2.9: AI model is explained, validated, and documented, *and* AI system output is interpreted within its context – *as identified in the **MAP** function* – ~~and~~ to inform responsible use and governance.
	MEASURE ~~2.9~~ 2.10: Privacy risk of the AI system – *as identified in the **MAP** function* – is examined regularly and documented.
	*MEASURE 2.11: Fairness and bias – as identified in the **MAP** function – are evaluated and results are documented.*
	MEASURE ~~2.10~~ 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.*

NIST AI 100-1
NIST AI Risk Management Framework
Measure Function Categories/Subcategories

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.

NIST AI 100-1
NIST AI Risk Management Framework
Measure Function Categories/Subcategories

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 AI lifecycle* are informed by *input from* domain experts and *relevant AI actors* ~~other stakeholder feedback~~ to validate whether the system is performing consistently as intended. Results are documented.
	MEASURE 4.3: Measurable performance improvements ~~(e.g., participatory methods) based on stakeholder consultations~~ *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.

NIST AI 100-1
NIST AI Risk Management Framework
Core Steps – 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 established in **GOVERN** and utilized in **MAP** and **MEASURE** bolster AI risk management efforts and increase transparency and accountability
- Processes for assessing emergent risks are in place, along with mechanisms for continual improvement

NIST AI 100-1
NIST AI Risk Management Framework
Manage Function Categories/Subcategories

MANAGE 1: AI risks based on impact assessments and other analytical output from the Map and Measure functions are prioritized, responded to, and managed.	MANAGE 1.1: *A* determination is made ~~about~~ *as to* whether the AI system achieves its intended purpose*s* and stated objectives and *whether its* development or deployment should proceed ~~in~~.
	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 most significant AI risks *deemed high priority*, *as* identified by the Map function, are developed, planned, and documented. Risk response options can include mitigating, transferring, ~~sharing,~~ 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.

NIST AI 100-1
NIST AI Risk Management Framework
Manage Function Categories/Subcategories

MANAGE 2: Strategies to maximize *AI* benefits and minimize negative impacts are planned, prepared, implemented, ~~and~~ documented, and informed by ~~stakeholder~~ 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~~,~~ - ~~and related reduction in severity of impact or~~ *to reduce the magnitude or* likelihood of ~~each~~ potential action*s*.
	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 ~~2.3~~ 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.

NIST AI 100-1
NIST AI Risk Management Framework
Manage Function Categories/Subcategories

MANAGE 3: *AI* Risks 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: ~~Responses to identified and measured risks~~ *Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks* are documented and monitored regularly.	MANAGE 4.1: Post-deployment system monitoring plans are implemented, including mechanisms for capturing and evaluating *input from* users and ~~stakeholder feedback~~ *other relevant AI actors*, appeal and override, decommissioning, incident response, *recovery,* and change management.
	MANAGE 4.2: Measurable *activities for* ~~continuous~~ *continual* improvement ~~activities~~ are integrated into *AI* system updates and include regular ~~stakeholder~~ 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

NIST AI 100-1
NIST AI Risk Management Framework
AI-Specific Risks

- The data used for building an AI system may not be a true or appropriate representation of the context or intended use of the AI system, and the ground truth may either not exist or not be available
- Harmful bias and other data quality issues can affect AI system trustworthiness, which could lead to negative impacts
- AI system dependency and reliance on data for training tasks, combined with increased volume and complexity typically associated with such data
- Intentional or unintentional changes during training may fundamentally alter AI system performance
- Datasets used to train AI systems may become detached from their original and intended context or may become stale or outdated relative to deployment context
- AI system scale and complexity (many systems contain billions or even trillions of decision points) housed within more traditional software applications
- Use of pre-trained models that can advance research and improve performance can also increase levels of statistical uncertainty and cause issues with bias management, scientific validity, and reproducibility.
- Higher degree of difficulty in predicting failure modes for emergent properties of large-scale pre-trained models

NIST AI 100-1
NIST AI Risk Management Framework
AI-Specific Risks

- Privacy risk due to enhanced data aggregation capability for AI systems
- AI systems may require more frequent maintenance and triggers for conducting corrective maintenance due to data, model, or concept drift
- Increased opacity and concerns about reproducibility
- Underdeveloped software testing standards and inability to document AI-based practices to the standard expected of traditionally engineered software for all but the simplest of cases.
- Difficulty in performing regular AI-based software testing, or determining what to test, since AI systems are not subject to the same controls as traditional code development
- Computational costs for developing AI systems and their impact on the environment and planet
- Inability to predict or detect the side effects of AI-based systems beyond statistical measures
- **Social and ethical impact of the use of AI systems**