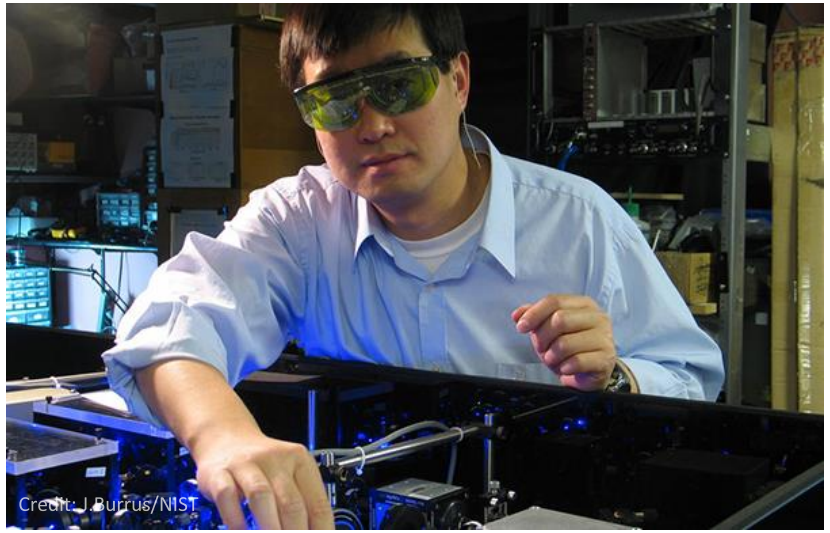
AI Risk Management Framework

National Institute of Standards and Technology
U.S. Department of Commerce

18 OCT 2022

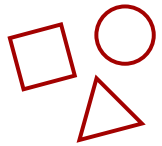
NIST's Mission

To promote U.S. innovation and industrial competitiveness by advancing **measurement science, standards,** and **technology** in ways that enhance economic security and improve our quality of life



Artificial Intelligence (AI) is rapidly transforming our world. New AI-enabled systems are revolutionizing and benefitting nearly all aspects of our society and economy – everything from commerce and healthcare to transportation and agriculture. But its development and use are not without challenges and risks.

NIST AI Program

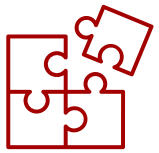


CONDUCT FOUNDATIONAL RESEARCH TO ADVANCE TRUSTWORTHY AI TECHNOLOGIES

ADVANCE AI RESEARCH AND INNOVATION ACROSS NIST'S LABORATORY PROGRAMS

ESTABLISH BENCHMARKS AND DEVELOP METRICS TO EVALUATE AI TECHNOLOGIES

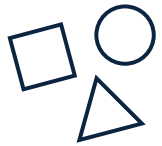
PARTICIPATE AND LEAD IN DEVELOPING STANDARDS TO ADVANCE AI INNOVATION



CONTRIBUTE NIST'S TECHNICAL EXPERTISE TO DISCUSSIONS AND DEVELOPMENT OF POLICIES

ENSURE NIST HAS RESOURCES AND EXPERTISE TO CARRY OUT ITS AI PROGRAMS

Key NIST Roles for the Federal Government



NIST AI RISK MANAGEMENT FRAMEWORK

NATIONAL AI ADVISORY COMMITTEE

AI RESEARCH RESOURCE TASK FORCE

FEDERAL AI STANDARDS COORDINATOR

INTERAGENCY COORDINATION WH OSTP/NSTC, TTC, QUAD

STAKEHOLDER OUTREACH

Trustworthy and Responsible AI @ NIST

Cultivate trust in the design, development, use and governance of artificial intelligence technologies and systems.

Development of AI Risk Management

AI Research, Standards and Evaluation

Establishing National AI Advisory Committee

AI RMF:
Managing Risk Through Trustworthy and Responsible AI

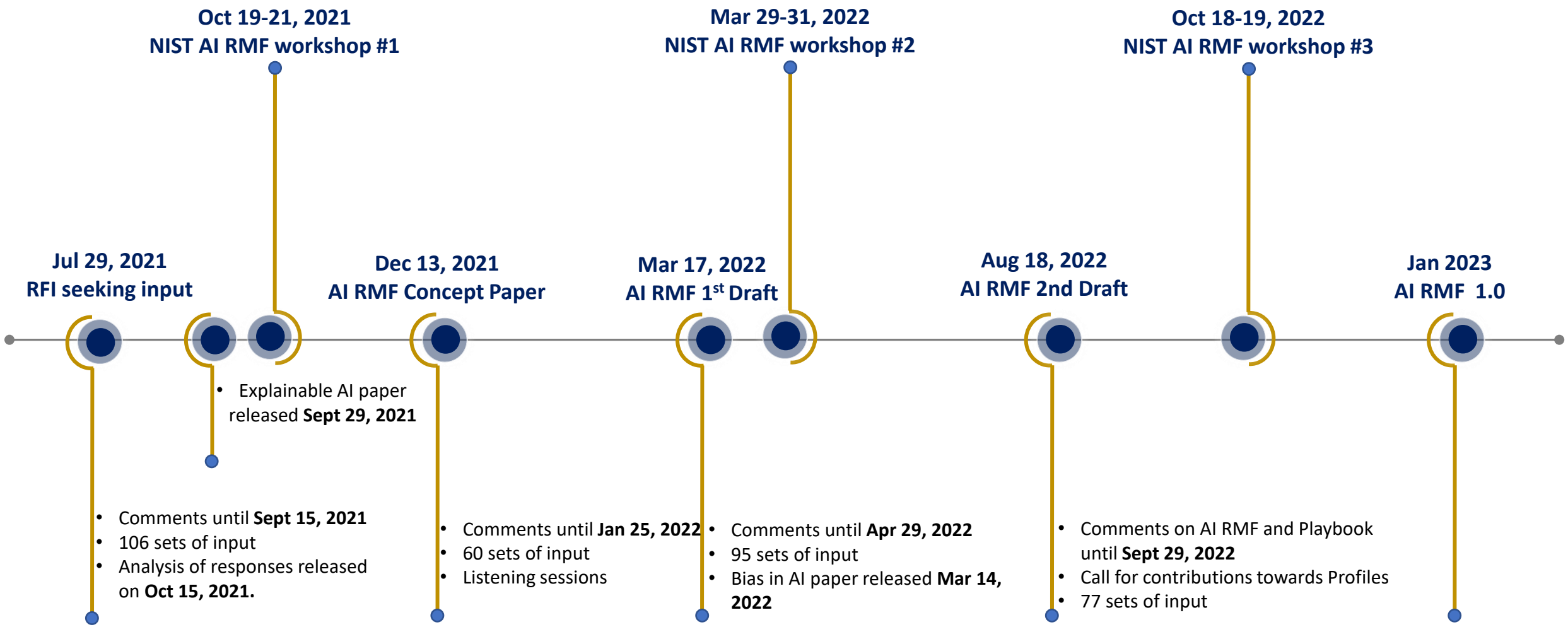
What

- Address risks to individuals, communities, organizations, and society
- Congressionally mandated, living document for voluntary use
- Maximize positive impacts, minimize potential negative impacts
- Rights-preserving, aims to operationalize values
- Law and regulation agnostic

How

- Developed in an open, transparent, collaborative process (ongoing)
- Outcome based
- Across context and use cases
- Trustworthy characteristics
- Responsible practices and culture (consideration of impacts)
- Inclusive and equitable

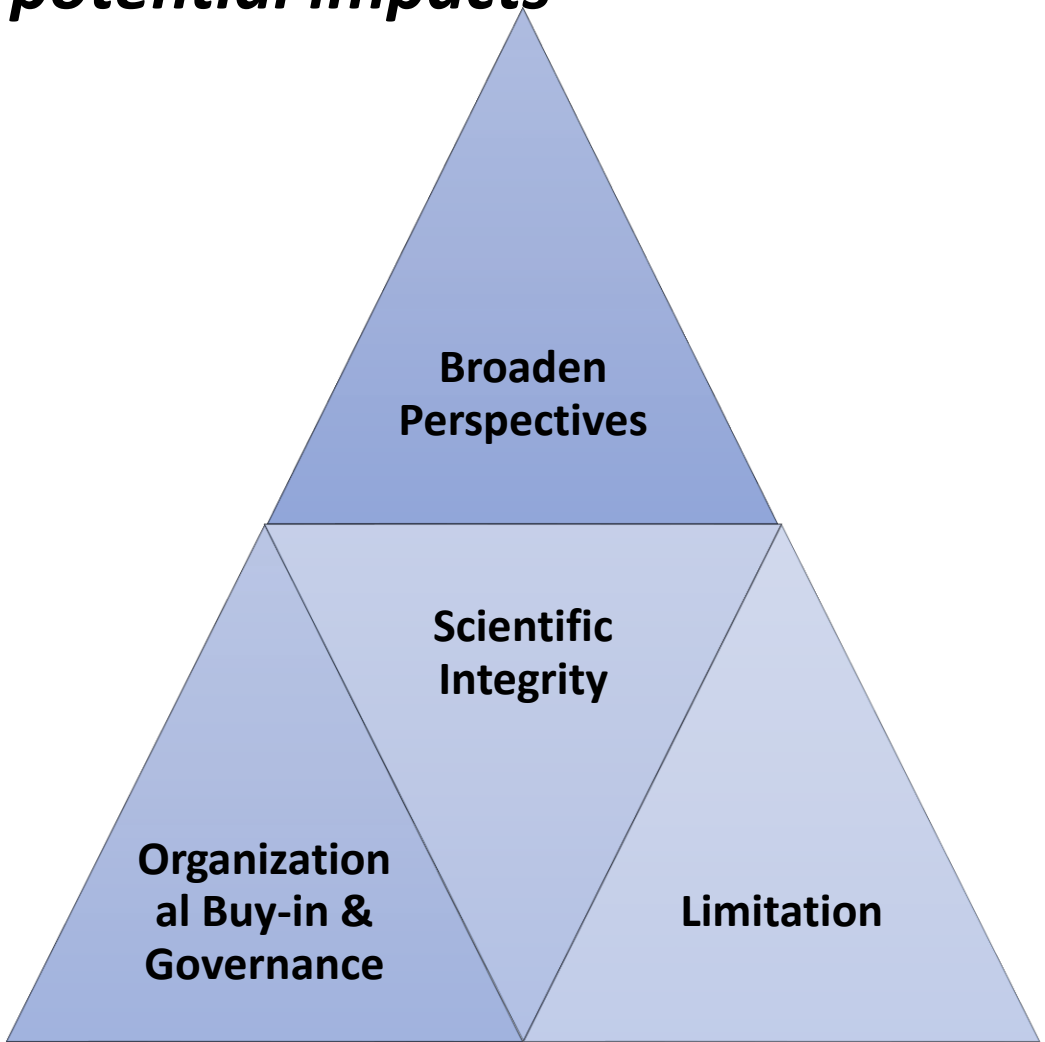
AI RMF Timeline and Engagements



Transforming Culture – Socio-technical Systems Approach

Takes into consideration the larger social context in which AI operates, its purpose and potential impacts

- Manage risk within/connected to specific operational **context**
  - utilize broader set of perspectives and expertise
  - apply **human-centered** design to AI systems
- Apply the **scientific method** to AI systems
- Set up **governance** structures for the people who build and maintain AI systems
- Consideration of **limitations** from an impact and values-based perspective



At A Glance


monitoring
measurement
managing
organizational
regularly
deployed
enhanced
subcategories
experts
individuals
security
stakeholder
negative
place
development
accountability
design
categories
managed
profiles
center
human
bias
potential
provide
existing
trustworthy
impacts
information
metrics
feedback
intended
framework
lifecycle
safety
decision
trustworthiness
governance
function
manage
practices
activities
guidance
training
identified
users
govern
outcomes
system
management
public
processes
factors
benefits
decisions
resources
identifying
deployment
lead
user
measuring
tevv
knowledge
tasks
actors
level
communities
organizations
context
understanding
category
risk
values
policies
model
responsible
measure
data
map
privacy
oecd
party
resource
technical
address
playbook
contexts
process
technology
society
models
stakeholders
groups
software
harms
transparent
documented
performance
characteristics
developed

Maximize Positive Impacts
Minimize Potential Negative Impacts

Risk refers to the composite measure of an event's probability of occurring and the consequences of the corresponding events. 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 tolerance refers to the organization's or stakeholder's readiness or appetite to bear the risk in order to achieve its objectives. Risk tolerance can be influenced by legal or regulatory requirements. While the AI RMF can be used to prioritize risk, *it does not prescribe risk tolerance*.

(Adapted from: ISO Guide 73).

Risk Management refers to coordinated activities to direct and control an organization with regard to risk.

(Source: ISO 31000:2018).

Risk Measurement (quantitively or qualitatively) is difficult, particularly for AI risks and impacts that are not well-defined or adequately understood.

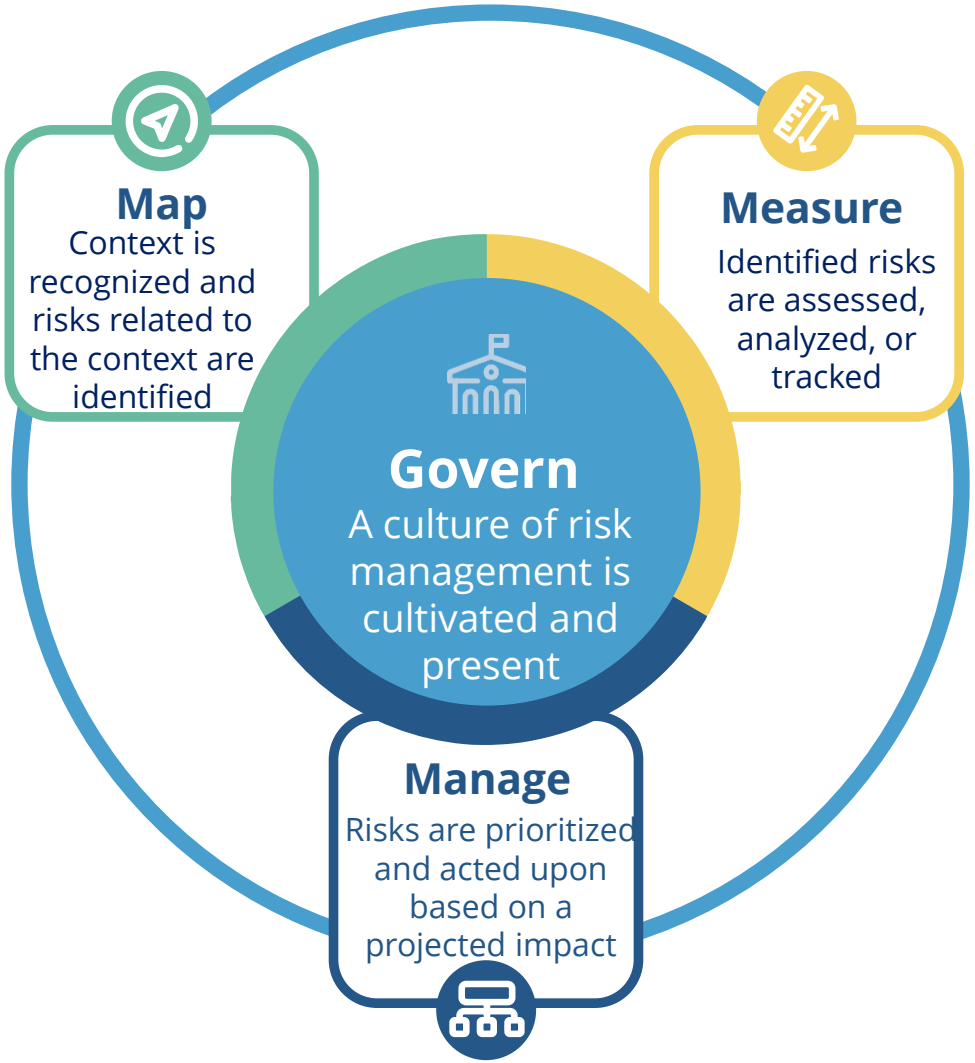
Audience

- AI Designers
- AI Developers
- AI Deployers
- AI Procurers
- AI Operators
- Third-party Entities
- Organizational Management, Senior Leadership
- End Users
- Affected Individuals/Communities

Trustworthy AI Characteristics

Trustworthy AI systems should achieve a high degree of control over risk while retaining a high level of performance quality. Achieving this difficult goal requires a comprehensive approach to risk management, with tradeoffs among the trustworthiness characteristics.

AI Risk Management Framework Core



Transforming Culture - Socio-technical approach takes into consideration the larger social system in which AI operates, its purpose and potential impacts

Categories in AI RMF GOVERN Function

GOVERN

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-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-3: Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle.

GOVERN-4: Organizational teams are committed to a culture that considers and communicates risk.

GOVERN-5: Processes are in place for robust stakeholder engagement.

GOVERN-6: Policies and procedures are in place to address AI risks arising from third-party software and data and other supply chain issues.

MAP

Categories in AI RMF MAP Function

MAP-1: Context is estabished and understood. +

MAP-2: Classification of the AI system is performed. +

MAP-3: AI capabilities, targeted usage, goals, and expected benefits and costs compared with the status quo are understood. +

MAP-4: Risks and benefits are mapped for third-party software and data. +

MAP-5: Impacts to individuals, groups, communities, organizational, or society are assessed. +

MEASURE

Categories in AI RMF MEASURE Function

MEASURE-1: Appropriate methods and metrics are identified and applied.

MEASURE-2: Systems are evaluated for trustworthy characteristics.

MEASURE-3: Mechanisms for tracking identified risks over time are in place.

MEASURE-4: Feedback about efficacy of measurement is gathered and assessed.

MANAGE

Categories in AI RMF MANAGE Function

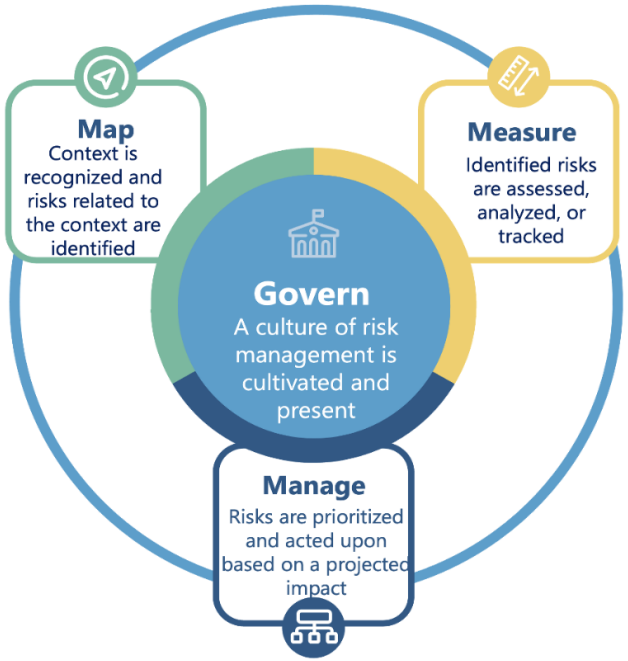
MANAGE-1: AI risks based on impact assessments and other analytical output from the Map and Measure functions are prioritized, responded to, and managed.

MANAGE-2: Strategies to maximize benefits and minimize negative impacts are planned, prepared, implemented, and documented, and informed by stakeholder input.

MANAGE-3: Risks from third-party entities are managed.

MANAGE-4: Responses to identified and measured risks are documented and monitored regularly.

NIST AI Risk Management Framework Playbook



Review FAQs

Welcome to the draft NIST AI Risk Management Framework (AI RMF) Playbook – a companion resource for the AI RMF.

The Playbook includes suggested actions, references, and documentation guidance for stakeholders to achieve the outcomes for "**Map**" and "**Govern**" – two of the four proposed functions in the AI RMF. Draft material for the other two functions, **Measure** and **Manage**, will be released at a later date.

This draft Playbook is being released to allow interested parties the opportunity to comment and contribute to the first complete version, to be released in January 2023 with the AI RMF 1.0. The Playbook is an online resource and will be hosted temporarily on GitHub Pages.

NIST welcomes feedback on this draft Playbook.

AI RMF Profiles

Use-case profiles

- Instantiations of the AI RMF functions, categories, and subcategories for a certain application or use case based on the requirements, risk tolerance, and resources of the Framework user.

Temporal profiles

- descriptions of either the current state or the desired, target state of specific AI risk management activities within a given sector, industry, organization, or application context

NIST welcomes contributions towards development of AI RMF use case profiles as well as current and target profiles.

Crosswalks

Table 1: Mapping of AI RMF taxonomy to AI policy documents.

AI RMF	OECD AI Recommendation	EU AI Act (Proposed)	EO 13960
Valid and reliable	Robustness	Technical robustness	Purposeful and performance driven Accurate, reliable, and effective Regularly monitored
Safe	Safety	Safety	Safe
Fair and bias is managed	Human-centered values and fairness	Non-discrimination Diversity and fairness Data governance	Lawful and respectful of our Nation's values
Secure and resilient	Security	Security & resilience	Secure and resilient
Transparent and accountable	Transparency and responsible disclosure Accountability	Transparency Accountability Human agency and oversight	Transparent Accountable Lawful and respectful of our Nation's values Responsible and traceable Regularly monitored
Explainable and interpretable	Explainability		Understandable by subject matter experts, users, and others, as appropriate
Privacy-enhanced	Human values; Respect for human rights	Privacy Data governance	Lawful and respectful of our Nation's values

NIST AI RMF Related Resources

AI RMF PLAYBOOK

AI RMF PROFILES

AI RMF GLOSSARY

AI STANDARDS HUB

AI METRICS HUB

…AND MORE

What's Next?

AI RMF Profile(s)

Work with SDOs on AI standards

Evaluations of AI RMF effectiveness

AI evaluations and Test beds

Trustworthy AI Resource Center

Crosswalks to other standards, frameworks, etc.

And more …

THANK YOU

Contact us via email at aiframework@nist.gov

For more info on the NIST AI RMF, visit https://www.nist.gov/itl/ai-risk-management-framework

24