AI Risk Management Framework
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
- Participate and lead in developing standards to advance AI innovation
- Contribute NIST’s technical expertise to discussions and development of policies
- Establish benchmarks and develop metrics to evaluate AI technologies
- 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**
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

- **Oct 19-21, 2021**
  - NIST AI RMF workshop #1
  - RFI seeking input
  - Comments until Sept 15, 2021
  - 106 sets of input

- **Dec 13, 2021**
  - AI RMF Concept Paper
  - Comments until Jan 25, 2022
  - 60 sets of input
  - Listening sessions

- **Mar 29-31, 2022**
  - NIST AI RMF workshop #2
  - AI RMF 1st Draft
  - Comments until Apr 29, 2022
  - 95 sets of input
  - Bias in AI paper released Mar 14, 2022

- **Aug 18, 2022**
  - AI RMF 2nd Draft
  - Comments on AI RMF and Playbook until Sept 29, 2022
  - Call for contributions towards Profiles
  - 77 sets of input

- **Jan 2023**
  - AI RMF 1.0

- **Oct 18-19, 2022**
  - NIST AI RMF workshop #3

**Explainable AI paper released Sept 29, 2021**
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
**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 Management** refers to coordinated activities to direct and control an organization with regard to risk.


**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 Measurement** (quantitively or qualitatively) is difficult, particularly for AI risks and impacts that are not well-defined or adequately understood.
<table>
<thead>
<tr>
<th>Audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI Designers</td>
</tr>
<tr>
<td>AI Developers</td>
</tr>
<tr>
<td>AI Deployers</td>
</tr>
<tr>
<td>AI Procurers</td>
</tr>
<tr>
<td>AI Operators</td>
</tr>
<tr>
<td>Third-party Entities</td>
</tr>
<tr>
<td>Organizational Management, Senior Leadership</td>
</tr>
<tr>
<td>End Users</td>
</tr>
<tr>
<td>Affected Individuals/Communities</td>
</tr>
</tbody>
</table>
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

Map
Context is recognized and risks related to the context are identified

Govern
A culture of risk management is cultivated and present

Measure
Identified risks are assessed, analyzed, or tracked

Manage
Risks are prioritized and acted upon based on a projected impact

Transforming Culture - Socio-technical approach takes into consideration the larger social system in which AI operates, its purpose and potential impacts
| 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. |
Categories in AI RMF MAP Function

MAP-1: Context is established 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 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.
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.
<table>
<thead>
<tr>
<th>AI RMF</th>
<th>OECD AI Recommendation</th>
<th>EU AI Act (Proposed)</th>
<th>EO 13960</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid and reliable</td>
<td>Robustness</td>
<td>Technical robustness</td>
<td>Purposeful and performance driven</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accurate, reliable, and effective</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Regularly monitored</td>
</tr>
<tr>
<td>Safe</td>
<td>Safety</td>
<td>Safety</td>
<td>Safe</td>
</tr>
<tr>
<td>Fair and bias is managed</td>
<td>Human-centered values and fairness</td>
<td>Non-discrimination</td>
<td>Lawful and respectful of our</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diversity and fairness</td>
<td>Nation’s values</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data governance</td>
<td></td>
</tr>
<tr>
<td>Secure and resilient</td>
<td>Security</td>
<td>Security &amp; resilience</td>
<td>Secure and resilient</td>
</tr>
<tr>
<td>Transparent and accountable</td>
<td>Transparency and responsible disclosure</td>
<td>Transparency</td>
<td>Transparent</td>
</tr>
<tr>
<td></td>
<td>Accountability</td>
<td>Accountability</td>
<td>Accountable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human agency and oversight</td>
<td>Lawful and respectful of our</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nation’s values</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Responsible and traceable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Regularly monitored</td>
</tr>
<tr>
<td>Explainable and interpretable</td>
<td>Explainability</td>
<td></td>
<td>Understandable by subject</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>matter experts, users, and others, as</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>appropriate</td>
</tr>
<tr>
<td>Privacy-enhanced</td>
<td>Human values; Respect for human rights</td>
<td>Privacy</td>
<td>Lawful and respectful of our</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data governance</td>
<td>Nation’s values</td>
</tr>
</tbody>
</table>
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