Artificial Intelligence Risk Management Framework (AI RMF 1.0)
WHAT IS THE AI RMF?

Voluntary resource for organizations designing, developing, deploying, or using AI systems to manage AI risks and promote trustworthy and responsible AI

- Rights-preserving
- Flexibly applied
- Measurable
THE PATH TO AI RMF 1.0

Oct 19-21, 2021
NIST AI RMF workshop #1

Dec 13, 2021
AI RMF Concept Paper
- Explainable AI paper released Sept 29, 2021
- Comments until Sept 15, 2021
- 106 sets of input

Mar 17, 2022
AI RMF 1st Draft
- Comments until Jan 25, 2022
- 59 sets of input
- Listening sessions
- Comments until Apr 29, 2022
- 92 sets of input
- Bias in AI paper released Mar 14, 2022

Mar 29-31, 2022
NIST AI RMF workshop #2

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

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

Jan 26, 2023
AI RMF 1.0
AI RMF Playbook
• **AI systems**: engineered or machine-based system that *generates outputs such as predictions, recommendations, or decisions* influencing real or virtual environments and operating with varying levels of autonomy.

• **Risk**: 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.
AI RISKS AND TRUSTWORTHINESS

- Safe
- Secure & Resilient
- Explainable & Interpretable
- Privacy-Enhanced
- Fair - With Harmful Bias Managed
- Accountable & Transparent

Valid & Reliable
AI RISK MANAGEMENT CHALLENGES

Risk measurement
Risk tolerance
Risk prioritization
Risk integration & management
AUDIENCE: AI LIFECYCLE AND AI ACTORS
AI RMF CORE

Map
Context is recognized and risks related to 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
NIST AI RMF Playbook

The Playbook provides suggested actions for achieving the outcomes laid out in the AI Risk Management Framework (AI RMF) Core (Tables 1–4 in AI RMF 1.0). Suggestions are aligned to each sub-category within the four AI RMF functions (Govern, Map, Measure, Manage).

The Playbook is neither a checklist nor set of steps to be followed in its entirety.

Playbook suggestions are voluntary. Organizations may utilize this information by borrowing as many—or as few—suggestions as apply to their industry use case or interests.

Download the NIST AI RMF Playbook

- Playbook PDF
- Playbook CSV
- Playbook Excel
- Playbook JSON
Implementations of the AI RMF functions, categories, and subcategories for a specific setting or application based on the requirements, risk tolerance, and resources of the Framework user.

- **Use-case profiles**: e.g., hiring or fair housing
- **Temporal profiles**: e.g., current state vs. the target state
- **Cross-sectoral profiles**: e.g., large language models, cloud-based services or acquisition
WHAT’S NEXT?

1. Alignment with international standards & production crosslinks to related standards.

2. Expanded TEVV efforts.

3. AI RMF 1.0 Profiles.

4. Guidance on the tradeoffs and relationships that may exist among trustworthiness characteristics.

5. Measuring the effectiveness of the AI RMF.


8. Guidance related to explainability and interpretability, and how to apply that guidance within the AI RMF.


10. Tutorials and other resources to enhance multi-disciplinary and social-technical approaches to AI risk management.
NIST TRUSTWORTHY AND RESPONSIBLE AI RESOURCE CENTER

https://airc.nist.gov/

AI RMF PLAYBOOK
AI RMF PROFILES
AI RISK GLOSSARY
AI STANDARDS TRACKER
AI METRICS HUB

...AND MORE

https://airc.nist.gov/

AIframework@nist.gov