https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai

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

Information Technology Laboratory (https://www.nist.gov/itl)

AI Risk Management Framework (https://www.nist.gov/itl/ai-risk-management-framework)

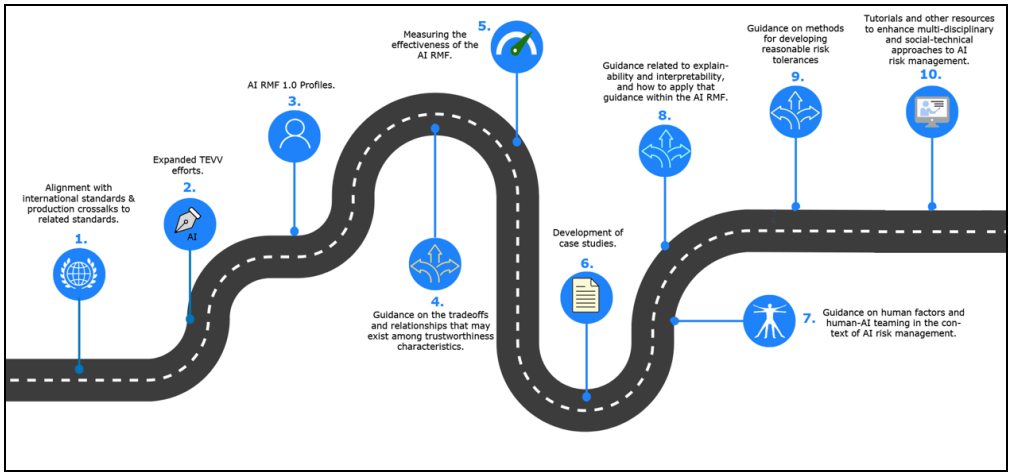
Roadmap for the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)

This Roadmap is a companion to the Artificial Intelligence Risk Management Framework (AI RMF 1.0). As directed by the National Artificial Intelligence Initiative Act of 2020 (P.L. 116-283), the goal of the AI RMF is to offer a *voluntary* 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.

The Framework 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. The AI RMF is designed to equip organizations and individuals – AI actors – with approaches that increase the trustworthiness of AI systems.

This Roadmap identifies key activities for advancing the AI RMF that *could be carried out by NIST in collaboration with private and public sector organizations – or by those organizations independently*. NIST's involvement will depend in part on resources available. Work described in the Roadmap is intended to help fill gaps in knowledge, practice, or guidance and be useful to a broader audience in pursuit of trustworthy and responsible AI. Roadmap activities will change as AI technologies and experience evolve.

Comments on this Roadmap are welcomed by NIST at any time and may refer to specific items that are either missing or incomplete, or express commitments to pursue Roadmap items. Comments should be addressed to AIframework@nist.gov (https://www.nist.govmailto:AIframework@nist.gov).



Areas that NIST considers to be top priorities are displayed In the graphic above and explained in the text below:

1. **Alignment with international standards and production crosswalks to related standards**. (e.g., ISO/IEC 5338, ISO/IEC 38507, ISO/IEC 22989, ISO/IEC 24028, ISO/IEC DIS 42001, and ISO/IEC NP 42005.)

   In its role as federal AI standards coordinator, NIST works across the government and with industry stakeholders to identify critical standards development activities, strategies, and gaps. Based on priorities outlined in the NIST-developed "Plan for Federal Engagement in AI Standards and Related Tools (https://www.nist.gov/document/report-plan-federal-engagement-developing-technical-standards-and-related-tools)," NIST is tracking AI standards development opportunities, periodically collecting and analyzing information about agencies' AI standards-related priority activities, and making recommendations through the interagency process to optimize engagement. NIST also participates in selected AI standards activities. See: https://www.nist.gov/artificial-intelligence/technical-ai-standards (https://www.nist.gov/artificial-intelligence/technical-ai-standards)

2. **Expanded TEVV efforts.** NIST will work with the broader community to develop tools, benchmarks, testbeds, and standardized methodologies for evaluating risks in AI and system trustworthiness, including from a socio-technical lens.

3. **AI RMF 1.0 Profiles**. Profiles are a primary method for organizations to share real-life examples of how they put the AI RMF into practice. These profiles can

be for a given industry sector (such as hiring, criminal justice, lending), cross-sectoral (such as large language models, cloud-based services or acquisition), temporal (such as current vs desired state), or other topics. Organizations and individuals are encouraged to jointly or independently produce AI RMF profiles.

4. **Guidance on the tradeoffs and relationships that may exist among trustworthiness characteristics**. A key challenge in managing AI risks is the need to balance and navigate tradeoffs between trustworthiness characteristics, and to do so systematically and dependent on the values at play in the variety of contexts and use cases. NIST intends to investigate key concepts in this topic area and develop guidance for navigating tradeoffs with the broader community.

5. **Measuring the effectiveness of the AI RMF.** As the AI RMF is adopted, organizations will have different experiences applying the Framework. These lessons and insights can improve the utility and effectiveness of the AI RMF and enhance our understanding of managing AI risks more generally. NIST will collaborate with experts in program evaluation, the AI RMF user community, and other interested parties to establish methods to capture, evaluate, and share insights about the Framework's use. These efforts will contribute to revisions to AI RMF 1.0 and future versions of this Roadmap.

6. **Development of case studies** about how the AI RMF has been used by a single organization or sector, context, or AI actor. Use cases may be similar to AI RMF profiles but can more directly describe organizational experiences and challenges using the AI RMF – and how they were addressed – along with information about resources, timeframes, and AI RMF effectiveness.

7. **Guidance on human factors and human-AI teaming in the context of AI risk management.** NIST intends to investigate how human-AI teams should best be configured to reduce likelihood of negative impacts/harms to individuals, groups, communities, and society. These impacts may be related to human *insights* about AI-produced output, or human *oversight* of AI systems and their operation in real world environments.

8. **Guidance related to explainability and interpretability, and how to apply that guidance within the AI RMF.** As our knowledge advances about how AI is understood across different audiences and for different purposes, NIST has identified a need to connect the field of AI explainability and interpretability more directly to AI risk management.

9. **Guidance on methods for developing reasonable risk tolerances.** NIST will work with the community to identify approaches organizations can use to develop risk tolerances.

10. **Tutorials and other resources to enhance multi-disciplinary and socio-technical approaches to AI risk management.** Topics associated with AI risk management are complex and evolving, and the community of interest will continue to expand and become increasingly multidisciplinary. To keep the broader AI community informed, NIST will encourage subject matter experts and other interested parties to develop education materials targeted to different audience types and will contribute its technical expertise to these efforts.

Created January 24, 2023, Updated March 14, 2023