AI Risk Management Framework: Initial Draft

March 17, 2022

This initial draft of the Artificial Intelligence Risk Management Framework (AI RMF, or Framework) builds on the concept paper released in December 2021 and incorporates the feedback received. The AI RMF is intended for voluntary use in addressing risks in the design, development, use, and evaluation of AI products, services, and systems.

AI research and deployment is evolving rapidly. For that reason, the AI RMF and its companion documents will evolve over time. When AI RMF 1.0 is issued in January 2023, NIST, working with stakeholders, intends to have built out the remaining sections to reflect new knowledge, awareness, and practices.

Part I of the AI RMF sets the stage for why the AI RMF is important and explains its intended use and audience. Part II includes the AI RMF Core and Profiles. Part III includes a companion Practice Guide to assist in adopting the AI RMF.

That Practice Guide which will be released for comment includes additional examples and practices that can assist in using the AI RMF. The Guide will be part of a NIST AI Resource Center that is being established.

NIST welcomes feedback on this initial draft and the related Practice Guide to inform further development of the AI RMF. Comments may be provided at a workshop on March 29-31, 2022, and also are strongly encouraged to be shared via email. NIST will produce a second draft for comment, as well as host a third workshop, before publishing AI RMF 1.0 in January 2023. **Please send comments on this initial draft to AIframework@nist.gov by April 29, 2022**.

Comments are especially requested on:

1. Whether the AI RMF appropriately covers and addresses AI risks, including with the right level of specificity for various use cases.
2. Whether the AI RMF is flexible enough to serve as a continuing resource considering evolving technology and standards landscape.
3. Whether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks.
4. Whether the functions, categories, and subcategories are complete, appropriate, and clearly stated.
5. Whether the AI RMF is in alignment with or leverages other frameworks and standards such as those developed or being developed by IEEE or ISO/IEC SC42.
6. Whether the AI RMF is in alignment with existing practices, and broader risk management practices.
7. What might be missing from the AI RMF.
8. Whether the soon to be published draft companion document citing AI risk management practices is useful as a complementary resource and what practices or standards should be added.
9. Others?

Note: This first draft does not include Implementation Tiers as considered in the concept paper. Implementation Tiers may be added later if stakeholders consider them to be a helpful feature in the AI RMF. Comments are welcome.

Table of Contents

AI Risk Management Framework: Initial Draft - Part 1: Motivation

1 Overview

Remarkable surges in artificial intelligence (AI) capabilities have led to a wide range of innovations with the potential to benefit nearly all aspects of our society and economy – everything from commerce and healthcare to transportation and cybersecurity. AI systems are used for tasks such as informing and advising people and taking actions where they can have beneficial impact, such as safety and housing.

AI systems sometimes do not operate as intended because they are making inferences from patterns observed in data rather than a true understanding of what causes those patterns. Ensuring that these inferences are helpful and not harmful in particular use cases – especially when inferences are rapidly scaled and amplified – is fundamental to trustworthy AI. While answers to the question of what makes an AI technology trustworthy differ, there are certain key characteristics which support trustworthiness, including accuracy, explainability and interpretability, privacy, reliability, robustness, safety, security (resilience) and mitigation of harmful bias. There also are key guiding principles to take into account such as accountability, fairness, and equity.

Cultivating trust and communication about how to understand and manage the risks of AI systems will help create opportunities for innovation and realize the full potential of this technology.

Many activities related to managing risk for AI are common to managing risk for other types of technology. An AI Risk Management Framework (AI RMF, or Framework) can address challenges unique to AI systems. This AI RMF is an initial attempt to describe how the risks from AI-based systems differ from other domains and to encourage and equip many different stakeholders in AI to address those risks purposefully.

It is important to note that the AI RMF is neither a checklist nor should be used in any way to certify an AI system. Likewise, using the AI RMF does not substitute for due diligence and judgment by organizations and individuals in deciding whether to design, develop, and deploy AI technologies – and if so, under what conditions.

This voluntary framework provides a flexible, structured, and measurable process to address AI risks throughout the AI lifecycle, offering guidance for the development and use of trustworthy and responsible AI. It is intended to improve understanding of and to help organizations manage both enterprise and societal risks related to the development, deployment, and use of AI systems. Adopting the AI RMF can assist organizations, industries, and society to understand and determine their acceptable levels of risk.

In addition, it can be used to map compliance considerations beyond those addressed by this framework, including existing regulations, laws, or other mandatory guidance.

Risks to any software or information-based system apply to AI; that includes important concerns related to cybersecurity, privacy, safety, and infrastructure. This framework aims to fill the gaps related specifically to AI. Rather than repeat information in other guidance, users of the AI RMF are encouraged to address those non-AI specific issues via guidance already available.

Part 1 of this framework establishes the context for the AI risk management process. Part 2 provides guidance on outcomes and activities to carry out that process to maximize the benefits and minimize the risks of AI. Part 3 [yet to be developed] assists in using the AI RMF and offers sample practices to be considered in carrying out this guidance, before, during, and after AI products, services, and systems are developed and deployed.

For the purposes of the NIST AI RMF the term *artificial intelligence* refers to algorithmic processes that learn from data in an automated or semi-automated manner.

The Framework, and supporting resources, will be updated and improved based on evolving technology and the standards landscape around the globe. In addition, as the AI RMF is put into use, additional lessons will be learned that can inform future updates and additional resources.

NIST's development of the AI RMF in collaboration with the private and public sectors is consistent with its broader AI efforts called for by the National AI Initiative Act of 2020 (P.L. 116-283), the National Security Commission on Artificial Intelligence recommendations, and the Plan for Federal Engagement in AI Standards and Related Tools. Engagement with the broad AI community during this Framework's development also informs AI research and development and evaluation by NIST and others.

2 Scope

The NIST AI RMF offers a process for managing risks related to AI systems across a wide spectrum of types, applications, and maturity. This framework is organized and intended to be understood and used by individuals and organizations, regardless of sector, size, or level of familiarity with a specific type of technology. Ultimately, it will be offered in multiple formats, including online versions, to provide maximum flexibility.

The AI RMF serves as a part of a broader NIST resource center containing documents, taxonomy, suggested toolkits, datasets, code, and other forms of technical guidance related to the development and implementation of trustworthy AI. Resources will include a knowledge base of terminology related to trustworthy and responsible AI and how those terms are used by different stakeholders.

The AI RMF is not a checklist nor a compliance mechanism to be used in isolation. It should be integrated within the organization developing and using AI and be incorporated into enterprise

Also available as part of the eCourse
[Responsible AI: Ethical Considerations and Building Trust through Transparency, Fairness, Accountability]

First appeared as part of the conference materials for the
35th Annual Technology Law Conference session
"Responsible AI: Ethical Considerations and Building Trust through Transparency, Fairness, Accountability"