MONITAUR

April 29, 2022

National Institute of Standards and Technology
Alicia Chambers, NIST Executive Secretariat
100 Bureau Drive
Gaithersburg, MD 20899

To whom it may concern,

We are pleased to have the opportunity to offer our feedback on the NIST Artificial Intelligence Risk Management Framework (AI RMF). This vital project has the potential to accelerate effective governance and assurance of artificial intelligence (AI) and machine learning (ML) systems.

At Monitaur, we believe that, by creating more trust and confidence in how these technologies are applied and managed, all stakeholders – corporations, regulators, and consumers – can benefit from extraordinary innovations that will improve our lives. We also believe that good AI requires great governance to ensure that these systems are more fair, safe, compliant, and robust than the human processes that they replace or enhance.

We recognize that NIST is at its core a technical organization seeking to provide clarity on the use of AI technologies, and the AI RMF achieves that aim. However, the risks associated with AI are not solely technical in nature, nor are we at a time in its maturity when we can mitigate those risks effectively with purely technical solutions. Recognizing those limitations, we encourage NIST to consider a holistic, lifecycle approach that incorporates oversight of the people and processes involved, in addition to the model and data risk management.

NIST previously delivered just such a comprehensive approach in its Cybersecurity Framework. In it, inherently technical activities (e.g. Detect) are complemented by human- and process-driven activities (e.g. Identify), as well as a recognition that technical activities must be supported by effective human effort. The combination of people, process, and technology enables organizations to mitigate risks, and we believe it should serve as a model for the AI RMF to create direction, clarity, and accountability for organizations that wish to use AI systems now and in the future.

In the below document, you will find our detailed responses listed by question. Thank you for your valuable time and consideration.

Sincerely,
Andrew Clark
Chief Technology Officer
Monitaur, Inc.

MONITAUR

Responses to RFI Questions

1. Whether the AI RMF appropriately covers and addresses AI risks, including with the right level of specificity for various use cases.

The break-up of technical characteristics and socio-technical characteristics required for a trustworthy system is well documented. The tangible, specific examples will help newer organizations with their AI RMF implementations.

NIST's Special Publication 1270 "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence " does a fantastic job of identifying different inputs to bias while also advocating for a holistic approach for remediation. However, this type of full lifecycle, holistic approach is less clearly described in the AI RMF. Anchoring around a full lifecycle approach (e.g. NIST's Cybersecurity Framework, CRISP-DM, or COBIT) ensures a broad-based approach to risk management, takes advantage of existing knowledge within organizations, and provides an accessible framework for newcomers.

Below are specific recommendations:

- Address risks related to the understanding of data, data governance, and general IT controls: We believe more traditional controls around IT systems, such as segregation of duties and access control, as well as controls around data provenance/governance, are an integral component of mitigating AI-related risks.
- Define full lifecycle governance and assurance more concretely: Identifying process steps and stakeholder roles in addition to technical characteristics will aid usage, compliance, and accountability since responsible individuals will have clarity on who is performing what actions.

2. Whether the AI RMF is flexible enough to serve as a continuing resource considering evolving technology and standards landscape.

See response to question #1.

3. Whether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks.

In this draft, NIST mentions three distinct frameworks:

- TEVV (Test, Evaluation, Validation, and Verification)
- Governance as "measure, manage, and map"

- Full lifecycle approach of pre-design, design and development, deployment, and test and evaluation

We expect organizations will struggle implementing the guidance of these three organizing principles, especially in areas with overlapping needs (e.g. “map” and “pre-design”). This could be simplified by organizing under a single holistic framework, such as the COBIT or NIST’s Cybersecurity Framework.

4. Whether the functions, categories, and subcategories are complete, appropriate, and clearly stated.

Many important aspects are present and communicated well in the AI RMF. Here are some examples of how NIST might incorporate elements from broader risk management frameworks:

- Although it is included in mapping section 6.1, we suggest highlighting how important it is for organizations to clearly establish that an AI solution is appropriate and best suited to the task for which it is designed. Documenting such a “Business Understanding” would prevent organizations from deploying AI to solve problems that are easily addressed with non-AI methods, and thereby reduce risk.
- The “Trustworthy AI: Risks & Characteristics” section is written with a technical slant, overlooking many non-technical risks of these systems.
- Even though accountability is included as a guiding principle, areas such as business applicability, data applicability, and other non-technical aspects are lacking within the framework.

With regard to the technical characteristics, we urge NIST to consider the following improvements:

- The AI RMF focus on explainability and interpretability prioritizes technical understanding over a broader understanding of how models make decisions. While explainability and interpretability are valuable tools for evaluating model predictions, without understanding the whole ML system, potential feedback loops or interdependencies may exist. To identify such critical problems, organizations should use full ML system scenario testing and transaction reperformance (ideally by objective reviewers).
- Concept drift, feature drift, and other applicable “model health” checks should be explicitly listed as important tools, particularly for medium- to high-risk models. The Safety section seems like a natural home for these and perhaps other technical risk management practices.

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.

NIST is farther along than most other regulatory bodies with regard to a risk management framework. Both IEEE and ISO/IEC SC42 are approaching the problem from a predominantly technical perspective as well. We believe that all 3 organizations would benefit from a more comprehensive approach with overall business implications and societal impact in mind. NIST's AI RMF should strive to go beyond IEEE and extend beyond the technical to encompass the broader world of risk.

6. Whether the AI RMF is in alignment with existing practices, and broader risk management practices.

The NIST AI RMF does not yet provide a clear link to existing risk management practices. Based on the document, NIST appears to have taken an "in to out" approach in developing the framework by focusing attention on terminology alignment and technical characteristics. Typically, risk management frameworks and best practices take an "out to in" approach instead, focusing on people and processes, the holistic drivers of risk, and execution throughout the stages of a project's entire lifecycle.

7. What might be missing from the AI RMF.

See response to question #4.

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.

A companion document providing detailed examples of AI risk management practices would be a very useful complement. This supplementary document would be most helpful and applicable if provided as an Excel spreadsheet – ideally with mapping to COBIT and NIST Cybersecurity frameworks.