600 14th St. NW, Suite 300
Washington, D.C. 20005

April 22, 2022

U.S. Department of Commerce
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

Subject: "AI Risk Management Framework: Initial Draft"

Dear Dr. Locascio:

On behalf of International Business Machines Corporation (IBM), we welcome the opportunity to respond to the National Institute of Standards and Technology's (NIST) request for information (RFI) regarding the "AI Risk Management Framework: Initial Draft" (hereafter, "draft AI RMF").

We appreciate NIST's work on this initiative, and believe this draft AI RMF is very much on the right trajectory. To help further advance this ongoing work, IBM offers three recommendations to improve this strong working draft:

- Clarify the definition of "risk" to ensure the "magnitude" of "adverse impacts" is taken into account;
- Incorporate the use of the terms "provider" and "owner" to further differentiate between those who develop AI systems and those who deploy them; and
- Segment risk considerations according to those immediately proximate to an organization's purview, and those risks that are society-wide.

IBM commends NIST for its work on the AI RMF and thanks you in advance for considering these comments. We welcome the opportunity to engage with the agency as it moves forward in this process.

Respectfully,

Chris Padilla
Vice President of Government and Regulatory Affairs
IBM Corporation

IBM Response to NIST RFI: AI Risk Management Framework

IBM appreciates the opportunity to respond to NIST's draft AI RMF. We have long been supportive of the agency's efforts in developing the AI RMF and we welcome the arrival of this draft framework. The comments that follow are organized according to the questions to which NIST has solicited answers.

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

The draft AI RMF does an excellent job of appropriately scoping AI risks, and we agree that the appropriate definition of risk is "a measure of the extent to which an entity is negatively influenced by a potential circumstance or event." Although we generally concur that these circumstances or events are primarily "a function of 1) the adverse impacts that could arise if the circumstance or event occurs; and 2) the likelihood of occurrence," **IBM suggests prefixing "magnitude" or "degree of" before "adverse impacts."** Risk is not merely the likelihood of *any* adverse impact resulting from an event, but the degree of the impact on the affected party.

Apart from this, we agree with the draft AI RMF's tripartite categorization of harms – harm to people, to organizations, and to systems – and NIST's decision to avoid prescribing risk thresholds or values. We particularly concur with the draft AI RMF's recognition that "risk tolerance – the level of risk or degree of uncertainty that is acceptable to organizations or society – is context and use-case specific," and that, as a result, "risk thresholds should be set through policies and norms that can be established by AI system owners, organizations, industries, communities, or regulators (who often are acting on behalf of individuals or societies)."

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

We appreciate NIST's close attention to the need for flexibility to be built into the draft AI RMF. Given ongoing developments at all levels of the AI field, this flexibility will ensure the framework's long-term resilience and viability as an authoritative and informative source of best practices. However, IBM believes **NIST could enhance the draft AI RMF's flexibility and longevity by further clarifying the distinction between the individual constituents described within the "AI system stakeholders" and "operators and evaluators" categories.**

The draft AI RMF defines AI system stakeholders as "those who have the most control and responsibility over the design, development, deployment, and acquisition of AI systems, and the implementation of AI risk management practices." This casts a large net, covering almost the entire breadth of the AI developmental lifecycle and supply chain, leaving little room for a more nuanced distinction between where various levels of accountability reside. Although it is true that there is a broadly shared responsibility for developing ethical AI systems, different stakeholders are better positioned to adopt different degrees – or types – of responsibility, and the expectations of accountability likewise differ based on where organizations and individuals exist in the larger AI ecosystem.

The definition of "operators and evaluators" begins to develop some of these distinctions, but also suffers from a lack of clarity. The terms include individuals who evaluate, validate, and verify system performance and explicitly points to "private sector researchers," "system operators," and "expert end users," among others. However, each of these individuals could also be interpreted as "AI system stakeholders" as well, creating additional confusion between the two categories.[1]

To provide further clarity regarding these terms, **IBM recommends the AI RMF use the terms "provider" and "owner" ("owner" being a term that, we would note, the draft AI RMF already makes use of in its discussion of risk thresholds) to further delineate between actors who develop AI systems and those who deploy them – and the expectations of accountability that accompany their scope of operational responsibility.**[2] That said, the Framework still must maintain its flexibility to accommodate that stakeholders do not always fit neatly into the audience categories and that risk mitigation responsibility should be allocated according to the role of the respective stakeholders in developing and deploying an AI system (e.g. curating training data sets, configuring AI systems, etc.).

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

---

[1] To take one example, a stakeholder who possesses "the most control and responsibility over the design … of AI systems" could easily be interpreted to be a "private sector researcher." This creates a lack of clarity regarding the distinction between the two stakeholder categories and where responsibilities and accountability for certain AI RMF processes would most appropriately reside.

[2] See Ryan Hagemann and Jean-Marc Leclerc, "Precision Regulation for Artificial Intelligence," *IBM Policy Lab*, 21 Jan. 2020, available at: https://www.ibm.com/blogs/policy/ai-precision-regulation/.

Although we agree that "a risk management framework should provide a structured, yet flexible" approach associated with the use of AI, the draft AI RMF draws a broad set of considerations that organizations are tasked with contemplating, from more immediate and knowable risks to broader "societal AI considerations and risks." To better segment the unique characteristics of these different portfolios of issues, **IBM recommends the draft AI RMF clarify that the considerations of risk fall into two distinct categories: (1) those risks within an organization's immediate purview of accountability, for which the RMF provides useful guidance in addressing, and (2) broader societal risks that lie outside the scope of the AI RMF, but which are nonetheless an important piece of the larger conversation surrounding the development and deployment of AI systems.**

This will have the benefit of focusing on clear, observable, and identifiable risks associated with discrete developments and deployments of AI, while tailoring appropriate processes and procedures to those stakeholders best positioned to address those risks.

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

One area where the draft AI RMF could improve in aligning its framework with real-world AI applications would be in further simplifying the AI RMF Core "functions-categories-subcategories" and "map-measure-manage" taxonomies. While we applaud NIST's attention to detail in this approach, we believe a design-development-deployment taxonomy – such as the one offered by *Confronting Bias; BSA's Framework to Build Trust in AI* – is more representative of existing approaches to identifying, managing, and mitigating risks posed by the development and deployment of AI systems.[3]

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.

IBM very much welcomes and supports that the draft AI RMF is generally aligned with existing approaches to AI standards development. The ongoing standardization work at the IEEE and ISO/IEC JTC 1 SC 42 reflect the state-of-the-art in delivering high quality standards addressing important aspects of AI. These include risk management standards, but also related topics, such as governance of

[3] See *Confronting Bias: BSA's Framework to Build Trust in AI*, BSA | The Software Alliance, 8 June 2021, pp. 19-27, available at: https://ai.bsa.org/wp-content/uploads/2021/06/2021bsaaibias.pdf.

AI systems, trustworthiness, transparency, preventing data bias, and more. IBM encourages NIST to build on and make use of these standards to leverage them for inclusion in the AI RMF.

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

As noted above, IBM believes the draft AI RMF is directionally aligned with many extant and still-developing standards and best practices. However, we believe there are a variety of opportunities to align the AI RMF Core with other existing best practices in risk management and recommend that NIST consider the *Confronting Bias* approach as a supplement to the broader AI RMF Core framework that aligns better with how organizations allocate responsibility for mitigating risk in AI systems.

7. What might be missing from the AI RMF.

As NIST continues its important work on the Guiding Principles, the RMF could benefit from acknowledging that AI risk management is a shared responsibility and that different entities will be responsible for different aspects of that risk management depending on their specific role in the AI digital ecosystem.

IBM believes that broad transparency with AI stakeholders is fundamental to mitigating AI risk and favors transparency over restrictions or overly prescriptive requirements. To that end, we recommend that NIST consider including in the transparency section of the Guiding Principles transparency about (1) when and whether users are interacting with AI, and (2) how the output of interactions between a user and an AI system will be used.

In addition, IBM would like to draw NIST's attention to ongoing developments in open source, in particular the Linux Foundation's LF AI & Data Project.[4] IBM recommends including these types of initiatives and activities within the scope of the AI RMF as many of these projects directly add to, or complement, technology development relevant for risk management and risk assessment.

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.

[4] See https://lfaidata.foundation/.

IBM supports the forthcoming draft companion and believes its inclusion will help further clarify risk management practices through illustrative case studies. We are also pleased to see the RMF "welcomes contributions of AI RMF profiles" to contribute to this companion document. IBM looks forward to offering such a contribution in the future.

Conclusion

IBM commends NIST for its tireless work in developing this draft AI RMF. We thank you for considering these comments and welcome the opportunity to engage with the agency as it moves forward in this process.