April 29, 2022

US Department of Commerce
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

Submitted via email: AIFramework@nist.gov

AI Risk Management Framework: Initial Draft
Comments of BSA | The Software Alliance

BSA | The Software Alliance appreciates the opportunity to provide feedback to the National Institute of Standards and Technology (NIST) regarding the Initial Draft of the AI Risk Management Framework (AI RMF).[1] BSA is the leading advocate for the global software industry before governments and in the international marketplace. Our members are at the forefront of software-enabled innovation that is fuelling global economic growth and helping businesses of all sizes leverage the benefits of cloud computing and AI-enabled products and services.[2] As leaders in AI development, BSA members have unique insights into both the tremendous potential that AI holds and the policies that can best support responsible AI innovation.

NIST's development of an AI RMF has the potential to serve as a much-needed foundation for global AI risk management efforts. By establishing a shared conceptual framework for identifying and mitigating risks throughout the AI system lifecycle, the AI RMF can serve as a common reference point that will facilitate communication between the many stakeholders involved in (or potentially impacted by) the development and deployment of an AI system. Given the global convergence around the need for risk-based regulatory approaches for AI, the AI RMF also has the potential to serve as an important tool for organizations to comply with emerging international legal requirements. More fundamentally, although the AI RMF is an inherently non-regulatory instrument, it can help

[1] AI Risk Management Framework: Initial Draft - March 17, 2022 (nist.gov)
[2] BSA's members include: Adobe, Alteryx, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, DocuSign, Dropbox, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, Prokon, PTC, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, Twilio, Unity Technologies, Inc., Workday, Zendesk, and Zoom Video Communications, Inc.

inform future regulatory efforts by providing a shared understanding about the multiple stakeholders, key considerations, and difficult trade-offs involved in effective risk management.

The Initial Draft represents a significant step toward achieving the full potential of the RMF. We are particularly encouraged by the Initial Draft's recognition that evaluating AI risk is an inherently context-specific inquiry that must account for the particular manner in which a system is (or will be) deployed. As the Initial Draft acknowledges, the risks implicated by an AI system, the appropriate measures for mitigating identified risks, and the appropriate thresholds for evaluating acceptable vs. unacceptable risk cannot be evaluated in the abstract. Instead, as the AI RMF recognizes, they must be informed by an understanding of the broader context in which the system will operate, including the "setting in which the AI system will be deployed," the business purpose it will be used for, and the "specific task" that it will support.

While effective risk management requires an informed understanding about how a system will be used, the groundwork for risk management must be laid far before a system is deployed. We therefore agree strongly with the Initial Draft that "risk management should be performed throughout the AI system lifecycle to ensure it is continuous and timely." In fact, the importance of a lifecycle-based approach to AI risk management prompted our recent effort to develop a framework for managing the risks of AI bias. The BSA Framework to Build Trust in AI (BSA Framework), outlines a methodology for performing impact assessments and corresponding best practices for mitigating risks that are pegged to the specific activities that occur at each stage of the AI lifecycle, including Design (i.e., project conception and data collection), Development (i.e., data preparation, model definition, validation, and testing), and Deployment.

Overall, the Initial Draft is built on a solid foundation. However, the ultimate success of the AI RMF will depend on ensuring that it is both flexible enough to accommodate the wide range of AI use cases, but specific enough so that it can serve as a useful mechanism to guide risk management practices. To assist NIST in striking that balance, we provide below a series of recommendations aimed at improving the usability of the framework and aligning it with industry best practices. In addition to the conceptual and structural recommendations below, we have attached an initial attempt to create a "crosswalk" comparison of the NIST AI RMF to the BSA Framework. We are pleased that the crosswalk demonstrates is a high degree of overlap between the BSA Framework and the Initial Draft. Included within the crosswalk are a number of specific recommendations for NIST's consideration that would help to fill in potential gaps in the AI RMF's categories and subcategories.

Recognizing that AI Risk Management is a Shared Responsibility

The Initial Draft would benefit from a discussion or guidance to help stakeholders navigate the complexities involved in AI risk management in circumstances where there may be multiple stakeholders involved in the development and deployment of an AI system. As currently drafted, the Framework Core in many ways seems to presuppose that the organization using the AI RMF will have full visibility into and control over the entire lifecycle of the AI system it is evaluating. For instance, while the Map Function largely focuses on an analysis of a system's deployment context,[3] it also calls for an examination of system artefacts that may only be available to the entity that trained the underlying model.[4]

The Framework Core provides little guidance about how organizations should utilize the RMF in circumstances where they are preparing to deploy an AI system or capability that may have been acquired from a third-party vendor.[5] Similarly, the Framework Core provides little guidance to vendors of general-purpose AI systems that may have limited insight into how their customers deploy the system.

NIST should consider adding high level guidance to Part 1 of the RMF to explain that risk management will in many instances be a shared responsibility that encompasses multiple entities, including organizations that develop AI systems for use by third parties (i.e., AI Developers) and entities that deploy AI systems that they have acquired from third parties (i.e., AI Deployers). While NIST should avoid drawing any bright lines about how specific responsibilities should be assigned, it would be helpful for the AI RMF to acknowledge that the appropriate allocation of risk management will depend on the nature of the underlying model and the extent to which it may be customized and/or re-trained by the AI Deployer.[6] Such guidance would serve as a useful tool for facilitating conversations between vendors of AI services and their customers to ensure that there is a shared understanding about their respective roles and responsibilities.

In addition, NIST should consider sub-dividing the "AI System Stakeholder" category in the Audience section of the Initial Draft. The Draft currently characterizes the AI System Stakeholder category 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." As noted above, not all System Stakeholders (as that category is currently defined) will have control over the full lifecycle of an AI system. NIST

[3] For example, the subcategories in the Map Function calls for an analysis of the "intended purpose [and] setting in which the AI system will be deployed," the "business purpose and context of use," and the "operation context in which the AI system will be deployed."

[4] For example, Map ID 2 includes a subcategory that is focused on "considerations related to data collection and selection." While it is not made explicit, it would appear that this subcategory is focused on *training* data.

[5] The Governance Function does acknowledge the importance of maintaining policies to "address AI risks arising from supply chain issues" including the "value and trustworthiness of third-party data or AI systems." But, beyond the mention of supply chain risk, the Framework Core lacks meaningful guidance about how to navigate these risks.

[6] The OECD for the Classification of AI Systems adopts a similar approach for assigning risk management responsibilities. See OECD Framework for the Classification of AI Systems, February 2022, https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1649808351&id=id&accname=guest&checksum=74B738F154B4F05D18B7B3D8B3477CE0, p. 48.

should therefore consider either teasing out “AI Developer” and “AI Deployer” organizations into separate (but potentially overlapping) categories[7] and/or acknowledging that not all “AI System Stakeholders” will have full control over an AI system’s lifecycle.

Finally, NIST should carefully scrutinize the use of the various AI System Stakeholder designations throughout the document to ensure that the RMF paints an accurate picture about their respective roles, responsibilities, and capabilities. We note for instance, that the discussion of “Technical Characteristics” on page 8 suggests that AI system accuracy, reliability, robustness, and resilience are “factors that are under the direct control of AI system designers and developers.” While it is true that AI system developers will in many circumstances have control over the accuracy of an AI system prior to its deployment, such control may be severed in instances when the system is deployed by a third-party. More generally, NIST should take care to avoid using AI stakeholder terminology in ways that may give rise to unintended interpretations or the appearance of policy preferences. We note for instance that the term “auditors” in the Operators & Evaluators Stakeholders grouping could be misconstrued as a reference only to third party organizations rather than in-house personnel. We urge NIST to clarify that the “auditors” referenced in the AI RMF can be personnel within an AI Developer or AI Deployer organization.

Provide Additional Guidance Regarding Use of the AI RMF Technical and Sociotechnical Characteristics and Clarify the Relationship to Impact Assessments[8]

The Initial Draft sets forth a risk assessment method that is centered around an analysis of a system’s <u>“technical” characteristics</u> (i.e., accuracy, reliability, robustness, resilience/ML security), <u>its “socio-technical” characteristics</u> (i.e., explainability, interpretability, privacy, safety, and managing bias), and a broader set of “<u>guiding principles</u>” (i.e., fairness, accountability, and transparency). For instance, the Map Function calls for:

- An analysis of the “potential business and societal (positive or adverse) impacts of technical and socio-technical characteristics for potential users, the organizations, or society”
- An elucidation of a system’s “potential harms…along technical and socio-technical characteristics” to ensure alignment with “guiding principles”

Similarly, the Measure Function calls for:

- An evaluation of “accuracy, reliability, robustness, resilience (or ML security), explainability and interpretability, privacy, safety, bias, and other system performance or assurance criteria.”

[7] See page 18 of the BSA Framework for a discussion of the relationship and definitions of key stakeholder groups: AI Developers, AI Deployers, and AI End-Users.

[8] This section includes recommendations that are responsive to NIST’s question about “[w]hether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks.”

We agree that an analysis of these system characteristics is an important element of effective risk management. However, the RMF provides little guidance about how these technical and sociotechnical characteristics should be used to guide the risk assessment. In the absence of additional explanation or guidance, we are concerned that characterizing the risk assessment process as centering on system characteristics may inadvertently suggest that AI risk management can be accomplished by looking only at attributes of the AI system in isolation from their impact on people. We recognize, of course, that this is not NIST's intent. Indeed, the Framing Risk section acknowledges that a core objective of AI RMF is to help organizations manage "enterprise and societal" risks and thereby prevent individual, organizational, and systemic harms.

We encourage NIST to provide additional clarification about how organizations should use the "technical and socio-technical characteristics" as heuristics for analyzing what impacts a system may have on internal and external stakeholders. It would be helpful, for instance, to include a sample analysis of a hypothetical AI system to demonstrate how "potential harms" can be "elucidated along technical and socio-technical characteristics and aligned with guiding principles" as contemplated by Map ID 4 Subcategory 2. Such guidance could be included in the "Practice Guide" that the Initial Draft mentions is currently under development.

As NIST considers how to structure potential additional guidance and/or adjustments, we recommend that it consider embracing the concept of "impact assessments" as a mechanism for assessing AI system risk. As noted in NIST Special Publication 1270, "impact assessments" are a "high-level structure that enables organizations to frame the risks of each algorithm or deployment while accounting for the specifics of each use case."[9] By clarifying that the risk analysis under the AI RMF should focus on the "impacts" of an AI system and providing more clarity about how the technical and socio-technical characteristics of AI systems should be used as part of that holistic examination, the Framework will better serve the needs of its target audience.

Focus on "Consequential" AI Systems

We recognize that the AI RMF is an explicitly voluntary resource that is intended to be flexible enough to accommodate systems of all kinds. However, given that the RMF is intended to serve as a resource for organizations of all sizes and "level[s] of familiarity" with AI, we urge NIST to consider adding in some high-level guidance to help organizations – particularly those who may be resource constrained – to identify AI systems within their purview that should be prioritized for assessment under the RMF.

[9] See NIST Special Publication 1270, Towards a Standard for Identifying and Managing Bias in Artificial Intelligence (nist.gov)

As you know, AI and AI-enabled features are now integrated into a staggering array of products and services that range from the trivial to the mission critical. Whether it is word processing software that uses ML-powered font recognition or email systems that utilize neural networks to detect spam, today's organizations may rely on an untold number of products and services that utilize some form of AI. As a practical matter it will be impossible for any organization to use the AI RMF to assess *every* AI-powered feature within an organization's supply chain. Unfortunately, the Initial Draft of the RMF currently lacks any guidance to help organizations identify the types of AI that should be prioritized for assessment via the RMF.

We recommend that NIST provide high level guidance to help organizations identify the "consequential" AI systems that should be prioritized for assessment under the RMF. Such guidance could highlight suggested criteria and/or questions for assessing whether a system is consequential. For instance:

- Is the system vital to an organization's ability operate?
- Is the system used in a manner that may produce legal effects on people and/or determine eligibility for important life opportunities?
- Could the system pose a risk of significant physical or psychological injury or otherwise threaten an individual's human rights?

* * * *

We appreciate the tremendous work that is reflected in the Initial Draft of the RMF and thank you for taking our recommendations into consideration.

Sincerely,

Christian Troncoso
Senior Director, Policy