January 25, 2022

Submitted via: AIframework@nist.gov

BSA | The Software Alliance appreciates the opportunity to provide feedback on the National Institute of Standards and Technology's (NIST) regarding the AI Risk Management Framework (AI RMF) Concept Paper.[1] BSA is an association of the world's leading enterprise software companies that provide businesses in every sector of the economy with tools to operate more competitively and innovate more responsibly.[2] As leaders in the development of enterprise AI, BSA members have unique insights into the technology's tremendous potential to spur digital transformation and the policies that can best support the responsible use of AI.

BSA has long recognized that risk management is a key component of promoting trust in AI and has strongly supported a NIST-led process to develop an AI risk management framework.[3] We are encouraged by the progress towards that goal that is reflected in the Concept Paper and appreciate the opportunity to provide initial feedback. Our feedback is informed by our recent experience working with BSA member companies to develop the BSA Framework to Build Trust in AI (the Framework),[4] a risk management framework for mitigating the potential for unintended bias throughout an AI system's lifecycle. Built on a vast body of

[1] https://www.nist.gov/news-events/news/2021/12/nist-seeks-comments-concept-paper-ai-risk-management-framework

[2] *BSA's members include: Adobe, Atlassian, Alteryx, Autodesk, Bentley Systems, Box, CNC/Mastercam, CrowdStrike, DocuSign, Dropbox, IBM, Informatica, MathWorks, Microsoft, Okta, Oracle, 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.*

[3] *See*, Comments of BSA | The Software Alliance Regarding NIST's Federal Artificial Intelligence Standards Engagement Plan (May 31, 2019), available at https://www.bsa.org/files/policy-filings/06102019bsasubmissionaistandardsrfi.pdf.

[4] Confronting Bias: The BSA Framework to Build Trust in AI (June 2021), available at https://ai.bsa.org/wp-content/uploads/2021/06/2021bsaaibias.pdf.

20 F Street, NW, Suite 800
Washington, DC 20001

P 202-872-5500
W bsa.org

research and informed by the experience of leading AI developers, the Framework outlines a lifecycle-based approach for performing impact assessments to identify risks of AI bias and highlights corresponding risk mitigation best practices.

We are pleased that the Concept Paper has proposed a structure for the AI RMF that closely resembles the approaches of the NIST Cybersecurity and Privacy frameworks. Organizing the AI RMF around a core composed of functions, categories, and subcategories is a sensible approach that will make it easier for organizations familiar with earlier NIST frameworks to more readily integrate the AI RMF into their broader risk management programs. The functions (Map, Measure, Manage, and Govern) outlined in the Concept Paper also provide a solid foundation that will help organizations adopt a lifecycle-based approach to managing the risks associated with AI systems that they are developing and/or using. To build on the solid foundation outlined in the Concept Paper, we offer below four recommendations for your consideration.

- ***Recommendation #1 – Building out the Categories for the Map Function***

We recognize that the categories identified in Table 1 of the Concept Paper are likely to be fleshed out. As you consider additions to the Map function, we encourage the inclusion of a category that addresses the key technical attributes and underlying components of an AI system that are integral to an examination of its potential risks. For instance, as discussed in the BSA Framework, a critical element of AI risk management involves a careful assessment of an AI system's "target variable" and its relationship to the data used to train the system.

A holistic understanding of what an AI system is designed to predict (i.e., its target variable) and the type of input data it relies on to make those predictions is essential for identifying (and managing) a broad range of risks that can emerge. The BSA Framework likewise highlights the important role that data provenance can play in identifying potential risks, including those that can emerge when training data is labeled and during the feature engineering process. Given the critical role that these sorts of system attributes can play in enumerating risks, we recommend that the Map function include an additional category that is focused on teasing out an AI system's core technical attributes and components.

NIST should also consider adding a category to the Map Function to help stakeholders identify when the risk profile of new applications of AI may differ from applications that have already been integrated into an organization's approach to risk management. Because many organizations may develop and/or deploy AI functionality with a consistent set of baseline risks, it may not be necessary to run a risk assessment entirely from scratch. Instead, the organization may integrate the new AI tool or application into a previously created Profile. To facilitate such activity, NIST should consider clarifying within the Map Function Category ID3 that organizations should seek to identify whether enumerated risks are unique to a particular application of AI or if they are risks that the organization has managed as part of earlier risk assessments.

- ***Recommendation #2 – Incorporating Impact Assessments in Govern Function***

We agree that "governance should be part of each function *and* a function of its own." As noted in the BSA Framework, effective risk management should be supported by a governance framework that sets forth the policies, processes, and personnel that an organization will use to perform impact assessments throughout the lifecycle of an AI system. Impact assessments complement organizational risk management by setting forth a framework for assessing the risk of individual AI applications. The BSA Framework, for instance, sets forth a methodology for performing impact assessments to identify and mitigate the risk of bias potentially associated with specific applications of AI. The AI RMF should acknowledge the role that impact assessments can play in overall AI risk management by including a category on impact assessments in the Govern Function.

- ***Recommendation #3 – Linking the Framework to Actual Risks***

Our third recommendation pertains to the Concept Paper's overall discussion of risk. While we strongly support the Concept Paper's *framing* of risk,[5] NIST should consider options for illustrating how the framework can be leveraged to manage

[5] For instance, we agree that the AI RMF should consider risk as a "composite measure of an event's probability of occurring and the consequences of the corresponding events." We likewise agree that "AI risk management is a as much about offering a path to minimize anticipated negative impacts of Ai systems, such as threats to civil liberties and rights, as it is about identifying opportunities to maximize positive impacts."

specific risks. We recognize that the goal of the AI RMF is to be a flexible resource that can be used by organizations of all types to address the myriad of risks that may be implicated by the broad range of AI applications. We support that approach, but it is possible that the level of abstraction at which risk is discussed in the Concept Paper could obscure the potential utility of the framework for addressing real-world concerns.

To avoid such an outcome, NIST should consider integrating the AI Risk Taxonomy[6] into the AI RMF and incorporating hypothetical examples to demonstrate how the framework can be used to map, measure, manage, and govern an array of specific AI risks. For instance, NIST could include hypothetical examples of organizations using the framework to develop Profiles for uses of AI that implicate the array of specific risks identified in the AI Risk Taxonomy.

- ***Recommendation #4 – Incorporating Metrics for Evaluating Overall Risk***

The AI RMF Measure Function proposes inclusion of metrics and other criteria for measuring specific risks identified and enumerated as part of the Map Function. It would be helpful for the AI RMF to also explore criteria that may be used to help stakeholders identify the overall risk level of an AI application, including criteria and methods that can be used to help characterize an AI application within a specific use case as "high-risk." This should not be an effort to try and categorize certain industries or uses as "high-risk," but instead identifying methods and criteria that stakeholders can use to help determine the level of aggregate risk for a particular use case.

[6] https://www.nist.gov/system/files/documents/2021/10/15/taxonomy_AI_risks.pdf