April 29, 2022

Unity welcomes the opportunity to provide input to the U.S. National Institute of Standards and Technology ("NIST") in its development of the Artificial Intelligence Risk Management Framework ("Framework" or "AI RMF"). Our comments center on three recommendations:

1. Provide guidance to identify AI systems that present greater risk
NIST should include guidance that enables users to identify systems that present greater risk, based on criteria outlined in the recently-published OECD Framework for the Classification of AI Systems.[1]

2. Elevate Technical and Socio-Technical Risks in the Framework Core
In the Framework Core, NIST should elevate the assessment of Technical and Socio-Technical Risks to the Category level and develop Subcategory guidance drawn from its framing narrative.

3. Illustrate shared responsibility models for AI systems in the Practice Guide
The forthcoming Practice Guide should provide examples of how Framework stakeholders (e.g., developers, end users, etc.) are typically best-positioned to help address them.

To enable NIST's analysis of stakeholder feedback, we have mapped our recommendations against NIST's proposed questions for the AI RMF stakeholder community in the footnotes. We look forward to further opportunities for engagement with NIST and the Framework stakeholder community on these recommendations and supporting comments

Recommendation 1: The Framework should provide guidance to identify AI systems that present greater risk

We believe that the Framework's practical utility would be enhanced by inclusion of criteria to identify systems that present greater inherent risk.[2] We note that the Initial Draft does not provide a risk level classification system, and we appreciate that it may not be possible to provide a universally-relevant tiered approach to risk. However, NIST should outline considerations that Framework users can apply in their shared evaluation processes. For example, when an AI developer at one organization is preparing a system for deployment by another actor, both parties would benefit from shared criteria to identify higher-risk AI systems. NIST should base these criteria on other leading proposals for risk analysis of AI systems. In

[1] OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en.
[2] This recommendation addresses questions 3, 5, and 7 presented by NIST in the Initial Draft.

the recently-published OECD Framework for the Classification of AI Systems, the Committee on Digital Economy Policy provided an analysis of "AI risk-based approaches to AI system application." This review looked across several leading risk classification proposals put forward by academic groups, governmental agencies and other actors, and standards development organizations, which identified three criteria that typically inform risk classification schemes.

- Scale (i.e., seriousness of adverse impacts (and probability))
- Scope (i.e., breadth of application, such as the number of individuals that are or will be affected)
- Optionality (i.e., degree of choice as to whether to be subject to the effects of an AI system)

NIST should include these criteria - or similar criteria - in the Framework Core in the Map function so that users incorporate them into their risk assessment activities. Expanding the Map function to encompass these criteria would reflect a logical extension of the Map function's current coverage. For example, the only reference to "scope" in the Initial Draft is in the Map function under 1D.3 ("Targeted application scope is specified and narrowed to the extent possible based on established context and AI system classification"). However, the Initial Draft does not plainly address scale nor optionality as contemplated by the OECD. These are highly relevant considerations, especially for assessment of Socio-Technical Characteristics.

Another benefit of these criteria is that they are agnostic about the use case under evaluation. They do not prescribe that a certain AI use case will be inherently higher-risk than another (e.g., use cases prohibited by or deemed High Risk by the proposed EU AI Act), but instead call for contextual assessment. Additionally, given that these criteria are drawn from leading risk classification initiatives being developed by institutions and organizations at the international and national level, these criteria should help the Framework remain broadly relevant even as new risk classification systems are published.

Recommendation 2: Elevate Technical and Socio-Technical Risks in the Framework Core
In the next iteration of the Framework Core, NIST should better connect its narrative descriptions from the AI Risks and Trustworthiness section with the Measure function.[3] The Initial Draft places all of the Socio-Technical and Technical Risk Characteristics into a single line under ID.2 of the Measure function, and it does not include any accompanying explanatory language on the page. This structure does a disservice to the importance of these considerations, and makes it difficult for users to identify and prioritize the risks that are most

[3] This recommendation addresses question 4 from the Initial Draft.

relevant to their assessment. At best, users would need to flip through the document to cross-reference uncommon terms like “explainability and interpretability” in the Framework Core with descriptions in the AI Risks and Trustworthiness narrative.

Specifically, NIST should identify assessment of Technical Risks and Socio-Technical Risks as two stand-alone Categories in the Measure function and develop related Subcategories based on the enumerated Risk Characteristics from the AI Risks and Trustworthiness section, including descriptive language from the underlying narrative. This approach would appropriately elevate these criteria in the AI RMF structure, and enhance ease of use. It would also strengthen the connection between the narrative portion of the AI RMF and the controls-oriented portion that will likely be reviewed most closely by target audiences.

Recommendation 3: Illustrate examples of shared responsibility for risk management in the Practice Guide

As a new domain of risk management, AI risk management would benefit from the shared responsibility model used by cloud services providers to allocate risk across organizational boundaries. During the rise of cloud computing, stakeholders grappled with a new model of outsourcing to shared infrastructure owned and operated by a third party. A key ingredient in cross-organizational dialogue was the development of a shared responsibility model, which set parameters for key parties (typically, the cloud service provider and their customer) about their responsibilities at different layers in the technical stack across typical cloud deployments (i.e., Software-, Platform-, or Infrastructure-as-a-Service).[4]

The Initial Draft demonstrates that addressing AI-related risks will also require effort from across organizational lines, but AI systems are not yet sufficiently standardized that “typical” systems can be leveraged to build models for shared responsibility that are broadly relevant. As it develops the Practice Guide, NIST should provide illustrative examples of how users shared responsibilities for risk management in trustworthy AI systems.[5] For example, these analyses could explore which parties (e.g., developer, deployer, etc.) took responsibility for certain risks based on the overall context surrounding the AI system. AI RMF users would benefit from documented examples of how organizations collaborated in this area, and what steps they took. These case studies could later inform a broadly-relevant shared responsibility model, but documenting current strategies would be especially helpful to organizations building their AI risk management programs today.

---

[4] Cloud Security Alliance (2020), “Shared Responsibility Model Explained”, https://cloudsecurityalliance.org/blog/2020/08/26/shared-responsibility-model-explained/.

[5] This recommendation addressed questions 1, 2, 6, and 8 from the Initial Draft.