Comments on NIST's Draft AI Risk Management Framework

Jacob Beswick, Director for AI Governance Solutions, Dataiku

April 2022

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

A. Reflecting on the structure and content of the AI RMF across the taxonomy, functions, categories and subcategories, it struck us that associating risks with 'stage of AI development' tends to be implied.

To test this reflection, we considered *unanticipated risks*, which may only be meaningfully identified upon their realization post-deployment. As an example, consider a situation where competing firms deploy pricing optimization models, and where such models, upon implementation, result in collusive pricing behaviors to the detriment of consumers. Such models may have been developed in alignment to the taxonomy, functions, categories and subcategories proposed. However, as in this scenario, it is possible to envision negative *impacts* in the post-deployment stage and to then consider that risks may arise in the (perhaps not-fully-understood) context of deployment and outside of the behaviors of a single company. This issue has been considered by the UK's CMA.

Currently, impacts are considered in the subcategories in relation to the Manage function. We would be interested to learn how the above scenario might play out within the AI RMF within the function, categories, and subcategories articulated.

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

A. The AI RMF's flexibility will be demonstrated over the course of future updates where these updates are informed by or reflect other substantial developments in the space of international technical standards and/or regulatory requirements at the horizontal and/or vertical level. As raised in the March workshop discussions, we would be keen to understand whether and how the AI RMF will coexist with other standards organizations' work in an effort to achieve harmony, rather than fragmentation.

B. It would be helpful to learn whether, and if so how, some components of the RMF will be fleshed out going forward. Related to flexibility, some areas, such as 'Organizational Integration', are light touch and from the perspective of an individual or small group of individuals looking to push forward an AI risk management agenda in an organization, may not be helpful beyond confirming the challenge that they are already experiencing. It would be a great benefit to extend the RMF - or the practical guidance - someday with AI RMF user stories, especially where very challenging aspects of AI risk management

are addressed, successfully overcome, or even that are unsuccessful. Doing so may contribute to meeting diverse audience needs while expanding key topics in the RMF.

C. See comment 5A.

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

A. Noting pages 1, 7, 18, and 19 that raise the challenges around implementing the Framework's proposals within an organization and team, finding the right balance between technological and sociological recommendations is challenging no matter the source. In our experience, often the sociological aspects of AI risk management and broader governance are moving, imperfect, and evolving. We agree with the comments on culture change and establishing roles and responsibilities, but would encourage you to emphasize that there is no one-size-fits-all approach nor is there likely to be many occasions of perfect-at-first-implementation. While this is the case, we would advocate that the balance between the key interventions and change management/culture change is lifted to a critical consideration.

B. We agree with the need to push AI risk management across organizations' management chains. We would question whether the document, as it stands, could sufficiently serve managerial personas who may not dive into the level of detail presented but require some motivation to think critically about AI risk management and the role of the framework. Building momentum across organizations can be extremely challenging and, starting at the operationally-focused part of the management chain could be for some organizations prohibitively complex.

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

A. The functions, categories, and subcategories section is in our reading the most impactful. While the introduction reads a little confusingly, the subsections, where each function is made applicable, helps to clarify.

B. There are uneven explicit references made to how the taxonomy's guiding principles are manifest in the functions, categories, and subcategories. For instance, unders 'systems are evaluated,' one would expect that the guiding principle of fairness is being observed but there is no explicit mention. The AI RMF is a meaningful extension of longstanding work on high concepts/principles insofar as it positions its users into the practicable. Therefore, a clearer, frank association between the taxonomy and the functions, categories, and subcategories would be quite helpful. This could be achieved simply by adding an additional column into the tables that clarifies which guiding principles are

being met by virtue of observing the categories and subcategories listed. Appreciating this could be clumsy on implementation, the direct link may nonetheless help some audiences.

C. With respect to 'Map-Context', we would advocate that, where reasonable, an accountable/responsible/owner person or team is clarified.

D. Under 'Map-Context', subject matter experts mentioned in 'Measure-Systems are evaluated' may provide early insight into the nature of risks associated with the context of application. Otherwise, risks may not be fully linked to socio-technical considerations.

E. Under 'Map-AI capabilities, targeted usage, etc.': assessing the cost of errors is the other side of the coin of ROI. Where ROI can be extremely difficult to map, we would expect cost of errors (or risks realized) would be equally challenging. As such, we would recommend adding an ordinal scale approach (high impact, some impact, little impact, no impact) or something to that effect. It relieves the burden of precision, which we have read into the 'Cost (monetary or otherwise)'.

F. Some aspects of 'Map' anticipate a linear process but, in practice, such linearity may not be universal. Systems may preexist use cases, optimized for those use cases/contexts post development. Considering or recognizing linear and nonlinear processes could be helpful for practicability.

G. Some of the above interventions may be mitigated if 'Govern' is brought up to be the first discussed in the list of functions. E.g. accountability structures. Similar to our company's approach, Governance starts with setting rules, requirements, and processes that reflect the priorities of the business. In the context of the AI RMF, these priorities are associated with identifying and mitigating risks. As such, it might flow that rules, requirements, and processes are established so that risks can be mapped, measured, and managed.

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.

A. The document commits to updating and improving the framework and supporting resources based on evolving technology and the global standards landscape. Appreciating the commitment to keep the framework up-to-date, it is possible that lag times between evolved expectations and the framework's update will exist. We would advocate that a notice of impending update be maintained to accommodate that lag period and to empower users to make informed decisions about whether and when to leverage the framework.

7. What might be missing from the AI RMF.

A. Having noted the association made between the AI RMF and it being used to map compliance considerations within existing regulations, law, etc., we would advocate that you identify regulators as a key stakeholder in Figure 1. Further to this, while the AI RMF is identified as a voluntary framework, we would strongly encourage you to impress on regulators that the RMF is designed to support AI developers and users with a way forward on risk management and, as such, may be considered a legitimate vehicle by regulators for compliance activities. While the USA will no doubt be your core audience, extending to other geographies preoccupied with these issues could facilitate future alignment in order to mitigate potential regulatory burdens for companies operating in multiple jurisdictions.

B. See 1A, 3A.

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. We view the companion document as critical to reinforcing the AI RMF's practicability. While the content in the AI RMF draft is useful, as discussed above, there are some areas where greater specification on operationalizing the functions, categories, and subcategories would be helpful.

B. With AI RMF users in mind, it could be helpful to *make clear where the AI RMF and companion documentation do not provide sufficient direction* and the rationale (e.g. beyond the scope). Without being explicit, users who treat these texts as necessary and sufficient may not progress on very challenging and fundamental areas. For example, the 'Organizational integration' and cultural change (under Govern) sections, which are fundamental to ensuring embedded and consistent operationalization of the framework.