Response of Microsoft Corporation to

NIST Artificial Intelligence Risk Management Framework Second Draft

September 29, 2022

Microsoft welcomes the opportunity to provide further feedback on the development of NIST's Draft AI Risk Management Framework (AI RMF). We are encouraged by the direction of the Framework and the way in which the latest draft of the AI RMF has taken on board feedback from Microsoft and others, including around the importance of addressing the sociotechnical nature of AI risk and the role impact assessments and testing of AI systems can play in guiding the risk identification and mitigation process. We set out thoughts for how we believe the AI RMF can be further strengthened and look forward to continuing to contribute to this process as NIST finalizes and launches the AI RMF.

Further guidance on utilizing impact assessments

Microsoft welcomes that the AI RMF Second Draft includes references to impact assessments. We view them as an important part of AI governance, providing a guided process by which to identify and mitigate the risks a system can pose, including to stakeholders across society. We encourage NIST to go further in helping organizations understand when they should use impact assessments to help implement the AI RMF, by including a new subcategory in the Govern function that recommends their use. This action could read: "An impact assessment for relevant AI systems is completed, documenting the applicable risks identified in the Map function and related mitigations." It would also be useful to include a short section articulating the value of impact assessments as part of the introduction to the AI RMF Core, highlighting the relevant subcategories that an impact assessment can help address. Microsoft recently shared its Impact Assessment template and Impact Assessment guide - developed to assist teams in completing a robust assessment – as part of the publication of our Responsible AI Standard. We share these materials in the hope they may help inform the ongoing discussion about building better norms and practices around AI.

Identification of responsibilities across the AI life cycle

It is helpful that the AI RMF sets out definitions of the different actors in the AI ecosystem, including those involved in developing AI systems as well as the entities deploying and monitoring their performance when in operation. To address the sociotechnical and context-specific nature of AI risks, it is important for both developers and deployers of AI systems to take responsibility at relevant points in the AI system life cycle. As NIST is aware, often the entity developing an AI system is different to the one deploying it. It will be helpful to clarify in Part 1 of the AI RMF that both developers and deployers will need to address elements of the Framework and the AI RMF could highlight the core elements that developers and deployers should incorporate into their internal processes. This would include outlining that both entities should map the potential impacts of a system on stakeholders across society and that both should test the system for appropriate performance. Those deploying a system should also offer training to individuals on how to use and oversee a system appropriately and be aware of the capabilities and limitations of the system in relation to their chosen use case. Developers should provide

this type of information on system performance and factors affecting use, to help deployers make informed deployment decisions.

Guidance around how to structure risk evaluation

The AI RMF sets out important steps that an organization can take to develop an overarching framework for identifying and addressing the risks AI systems can pose. An important element of this is helping organizations engage in structured risk evaluation, allowing them to quantify the types and level of risk a system may pose and calibrate related thresholds, alongside developing mitigations.

To do so, we believe the AI RMF should go further in setting out a baseline of key risk concepts and processes that organizations can use to advance a common internal foundation for risk evaluation. This is important given the many teams across an organization that will be involved in implementing the AI RMF. These include individuals involved in developing and deploying AI systems but also those responsible for enterprise risk management, corporate governance and compliance, privacy, cybersecurity, and more. The AI RMF defines the different characteristics of trustworthy AI, but a Framework user without robust AI knowledge may not be familiar with (1) how risks manifest from a failure to meet such characteristics, (2) how to weigh tradeoffs appropriately between competing characteristics, or (3) the types of mitigations that may be applied. Sharing more around these baseline concepts may be helpful to meet the needs of a multidisciplinary audience.

The AI RMF should also expand upon the types of factors that may increase an AI system's risk level and would benefit from setting out examples of how an organization may go about evaluating risk, including scenarios in which it is challenging to assess system risk in a precise fashion due to new and evolving sociotechnical risks.

In addition, we think it would be useful for NIST to highlight techniques, methodologies, and examples in the AI RMF Playbook that organizations can use when conducting risk evaluation. Currently, the Playbook links to frameworks on accountability and governance, and it would benefit from including practical and implementable tools in the next iteration—especially ones that assist with complex evaluations where AI spans software, hardware, and embedded solutions. Risk matrices, for example, are commonly used in the domain of system safety engineering to help assess large systems holistically. Sharing information about other approaches that organizations can take will be helpful, as will helping organizations understand where relevant work, including on related standards, is needed. We also wish to highlight sections 6.3.4-6.4.4 in ISO/IEC 23894 in Information technology — Artificial intelligence — Guidance on risk management, which is likely to soon be finalized. These sections may be helpful to the discussion around frameworks for evaluating qualitative and quantitative risks.

Advancing cross team collaboration in implementing the AI RMF

The AI RMF will benefit from illuminating how teams involved in implementing the Framework interact. We note the new language in Section 4.7 of the AI RMF Second Draft around how the AI RMF may interact with the Privacy Framework which is helpful and suggest that it may be beneficial to add additional guidance around how the Privacy and Cybersecurity Frameworks and AI RMF align. Visual cross domain mapping may be helpful. We highlight the following image that was in an initial draft of the NIST Privacy Framework. While this was not included in the finalized Privacy Framework, we feel this

type of visualization of responsibilities can be helpful in helping teams understand how their respective responsibilities overlap, both across the AI RMF and the interaction of the AI RMF with the other frameworks noted above. The Venn diagram on page 3 of the NIST Privacy Framework is also a helpful visual that could potentially be developed further with the addition of AI related risks.

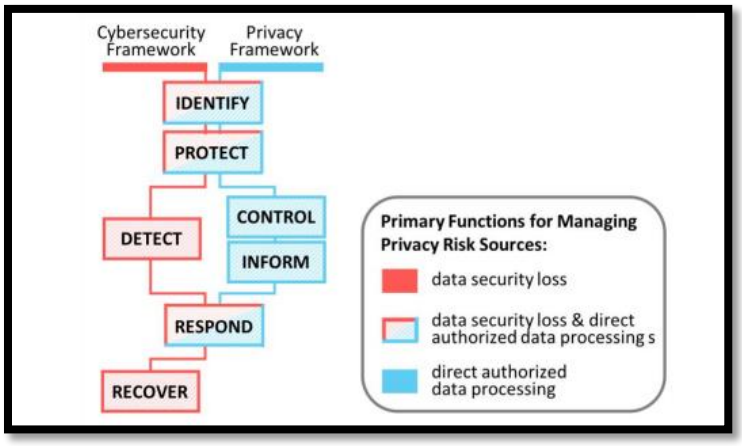

Figure 1: Cybersecurity Framework and Privacy Framework Functions Relationship

Aligning to international standards

We encourage NIST to align the AI RMF to international standards to help ensure compatibility with regulatory and governance frameworks across jurisdictions. This will be an ongoing process as the many standards currently in development in the responsible AI space continue to be formalized. It may be helpful to share mapping and/or guidance around relevant completed standards and how they apply to the Framework, including as part of future versions of the AI RMF. Specific standards that it will be important for the AI RMF to address include ISO/IEC FDIS 23894 – AI Risk Management and ISO/IEC 5338 – AI system life cycle processes, ISO/IEC 38507 — Governance Implications of the Use of Artificial Intelligence by Organizations, ISO/IEC 24028 – Overview of Trustworthiness in Artificial Intelligence. We also highlight ISO/IEC NP 42005 – AI System Impact Assessment. This is still at an early stage of development but will overlap with key elements of the AI RMF, including around mapping the impacts of an AI system on individuals across society. It will likely play an important role in helping organizations implement the AI RMF into the future.

We also believe there is an important role that NIST can play in continuing to engage in international standards conversations. We would encourage NIST to continue to share its work around the AI RMF to help develop international standards that can advance the practice of identifying and addressing AI risk.

Expanding governance responsibilities to strengthen oversight and accountability

The AI RMF's Govern function sets out important elements of an AI governance program. We encourage NIST to consider the inclusion of two additional subcategories to strengthen oversight and responsibility for relevant AI systems. First, we recommend a new Govern subcategory addressing system inventorying, along the lines of: "A mechanism is in place to inventory relevant AI systems and ensure appropriate oversight for systems involved in high-risk scenarios." There may be scenarios where it may

not be feasible or necessary to maintain an exhaustive inventory, for instance due to an AI system's integration into traditional software applications such as email spam filtering, however this type of inventorying is an important part of identifying and addressing AI risk. Organizations should establish their risk tolerance and determine which systems should be in scope.

Second, the Framework should elevate the importance of executive accountability as part of its governance function. We suggest a slight reframing to Govern 2.3 to read along the lines of: "Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment."

Future AI RMF roadmap and incorporation of TEVV and human factors work

It may be valuable to provide a roadmap for the AI RMF, setting out areas for further development and highlighting the way in which adjacent NIST research streams, for example the important ongoing work around human factors or testing, evaluation, verification and validation, will be incorporated into the AI RMF. Setting out a forward looking plan may be helpful in structuring collaboration and input from across different stakeholder groups, as well as in helping organizations understand how the AI RMF may develop.

Conclusion

Microsoft appreciates the opportunity to provide comments on the Second Draft of the NIST AI RMF. We believe that the framework will be an important tool for helping advance responsible AI practice and identify and address the risks that AI can pose. We look forward to continuing to contribute to this process as NIST finalizes and launches the AI RMF.