All comments will be made public as-is, with no edits or redactions. Please be careful to not include confidential business or personal information, otherwise sensitive or protected information, or any information you do not wish to be posted.

Comment Template for Responses to NIST Artifical Intelligence Risk Management Framework Request for Infromation (RFI)

Submit comments by August 19, 2021:

General RFI Topics (Use as many lines as you like)	**Response #**	**Responding organization**	**Responder's name**	**Paper Section (if applicable)**	**Response/Comment (Include rationale)**	**Suggested change**
Responses to Specific Request for information (pages 11,12, 13 and 14 of the RFI)						
1. The greatest challenges in improving how AI actors manage AI-related risks – where "manage" means identify, assess, prioritize, respond to, or communicate those risks;	1	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		which we can apply statistical and decision theoretic approaches to risk management. With AI Systems, both the system structure and system state are evolving, and the time constants on the dynamics of systems state and systems structure are different. All of that contributes to the complexity of AI systems. One of the greatest challenges is getting actors to see the whole system and hold the inherent complexity. Many want to approach AI systems and their risks linearly, tracking cause and effect. With AI, a necessary shift is to consider emergent issues and risks as components of interconnected and interacting systems rather than as independent issues with unrelated consequences. Addressing a risk likely means creating new vulnerabilities and new systems tradeoffs. Improvements in management of AI-related risks requires new approaches that reflect a whole systems perspective. As part of that, organizations need new approaches that broaden the scope of risk-based decisions to include opportunistic risk as well as possible threats. More generally, an AI system can only address risks that are known and within the purview of the system. An additional great challenge in improving the management of risk is then in making systems that can deal with complexity and which are built in ways that consider broad sets of risks. This	First, consider including response options of "enhance, exploit, and share" that can go along with the original response options provided that include, "avoid, mitigate, transfer, and accept". This will allow organizations to strike a balance between possible threats and opportunities. Currently, the question is posed for threats only. Second, consider including a direct focus on end user experience.

2. How organizations currently define and manage characteristics of AI trustworthiness and whether there are important characteristics which should be considered in the Framework besides: accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security (resilience), and mitigation of harmful bias, or harmful outcomes from misuse of the AI;	2	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		that a user places on the recommendations of an AI system in the human-AI setting (be it dyadic or team) which is determined as much by how robust the recommendations are, understandability and error rates (type 1 and 2) and the costs of errors and who bears it (think lawsuits and insurance). Another kind of trust is the community or individuals who are affected by the actions taken via the AI system and their assessment of whether the systems is trustworthy (think police and its use AI systems for face recognition). These characteristics of AI trustworthiness that are listed can be grouped in several ways. One grouping might be: - performance characteristics (accuracy), - deployment characteristics (reliability, robustness, safety, resilience), - adversarial characteristics (security, privacy, harmful outcomes from misuse of the AI), and - usability characteristics (explainability, interpretability, mitigation of harmful bias). Consider moving to a higher abstraction for the Framework to elicit the trust characteristics across a range of contexts; for example, for an Object Detection AI, mean Average Precision (mAP) is usually used instead of accuracy.	First, we suggest using a higher level of abstraction to guide the definition of trust characteristics for the Framework. For example, there are many additional performance characteristics, beyond accuracy, e.g. mAP, precision / recall, etc., and many additional deployment characteristics, e.g. uncertainty quantification, and so on. By moving to a higher level of abstraction, one may be better able to elicit the characteristics of trustworthiness relevant for the context of a given AI system. More broadly, consider not just the characteristics of trust in the AI system, but also the ability of the organization to build a trustworthy system. For example, does the organization have a documented risk appetite statement that enables standardization in risk decisions in adopting or building AI systems? If not, it may be unlikely that the organization can be trusted with its AI, even if the AI itself has many of the trust characteristics.
3. How organizations currently define and manage principles of AI trustworthiness and whether there are important principles which should be considered in the Framework besides: transparency, fairness, and accountability;	3	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		To implement principles like these, one will need to be able to measure them within the Framework, either at a quantitative level, similar to the characteristics of trustworthiness, or at a qualitative level. For example, accountability could be measured by organizations having a documented governance structure where accountability is chartered by role.	NIST may consider the use of a "Maturity Model-like" set of criteria to help organizations scale and adapt to properly account for the trustworthiness of AI and its use. This will allow for consistent qualitative measurement.
4. The extent to which AI risks are incorporated into different organizations' overarching enterprise risk management – including, but not limited to, the management of risks related to cybersecurity, privacy, and safety;	4	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		The connection to ERM is well stated. Note, however, that there are additional risks interdependencies. For example, talent recruitment and retention is another critical risk to consider given the degree of technical complexity and demands of AI.	We recommend the addition of other risk interdependencies here such as talent recruitment and retention. Additional ties of interconnectivity with an ERM portfolio could include strategy (e.g. mergers and acquisition), supply chain risk management (e.g. assessing the use of AI related product liability), and ethics (e.g ethical implementation of the technology). Furthermore, we suggest including guidance on what is similar and what needs to be different based on domain of application. For example, ERM in the context of AI for electricity grid anomaly detection is very different from risks for AI for
5. Standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles to identify, assess, prioritize, mitigate, or communicate AI risk and whether any currently meet the minimum attributes described above;	5	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		1. Explain and then speculate about the overall system - What problem are we solving and for whom? Is AI the right solution for the problem? Why? - Use a set of ethics to support the team in this work such as the - DoD's Principles for Ethical AI DOD Adopts Ethical Principles for Artificial Intelligence > U.S. Department of Defense > Release - Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities | U.S. GAO - Awesome AI Guidelines on GitHub from EthicalML: https://github.com/EthicalML/awesome-artificial-intelligence-guidelines - Consider what is interesting about this system to potential adversaries? - Consider what access adversaries might gain? What systems are connected? - Conduct speculative activities: - Checklist to prompt intentional, uncomfortable conversations: Designing Ethical AI Experiences (Carnegie Mellon University, Software Engineering Institute): https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=636620 - Harms modeling (Microsoft): https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/harms-modeling/ - Abusability Testing: UX in the Age of Abusability. The role of Composition,	

6. How current regulatory or regulatory reporting requirements (e.g., local, state, national, international) relate to the use of AI standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles;						
7. AI risk management standards, frameworks, models, methodologies, tools, guidelines and best practices, principles, and practices which NIST should consider to ensure that the AI RMF aligns with and supports other efforts;	6	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		We concur with the ERM-based approach for risk management here, as it recognizes the interdependency of AI related risks with others in the ERM risk portfolio.	An additional document to assist NIST and its readers in the development of ERM policies and practices could include SEI's OCTAVE FORTE. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=644636
8. How organizations take into account benefits and issues related to inclusiveness in AI design, development, use and evaluation – and how AI design and development may be carried out in a way that reduces or manages the risk of potential negative impact on individuals, groups, and society.	7	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		AI systems learn from examples, so it helps to have a diverse team that can bring different lenses to a problem and identify appropriate datasets to train the AI system on. It naturally follows that assembling a team with different backgrounds that can speak to different aspects of the problem will result in a better selection of datasets. AI teams need to be informed by a range of cultures, experiences, and how team members think about the world and the heuristics they use to solve problems. A team can be made up of members with diverse backgrounds, but if all the team members are engineers, they will approach the problem space in the same way. Teams need to explore what it would mean to partner with a policy maker or a philosopher and how those unique perspectives would drive solutions that would be ethical and implementable.	
9. The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes");						
10. Effective ways to structure the Framework to achieve the desired goals, including, but not limited to, integrating AI risk management processes with organizational processes for developing products and services for better outcomes in terms of trustworthiness and management of AI risks. Respondents are asked to identify any current models which would be effective. These could include – but are not limited to – the NIST Cybersecurity Framework or Privacy Framework, which focus on outcomes, functions, categories and subcategories and also offer options for developing profiles reflecting current and desired approaches as well as tiers to describe degree of framework implementation; and	8	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		Positioning the framework as a continuous learning process (see, for example Kolb's experiential learning model) can help to introduce the notion that everyone has a role to play in learning about the evolution of AI systems, the risks that emerge, and strategies for addressing them. By focusing the framework on learning toward the desired systems outcomes (i.e., systems that are trustworthy, secure, resilient, etc.) it broadens the aperture to include multiple approaches for how to reach end states, rather than focus on a single approach adopted by individuals with fixed roles and skills. Additionally, change management will be a critical part of adopting new risk management approaches as AI systems have several inherent differences from traditional software risk management and thus, Kotter's change management model might also prove useful.	

11. How the Framework could be developed to advance the recruitment, hiring, development, and retention of a knowledgeable and skilled workforce necessary to perform AI-related functions within organizations.	9	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		An assumption often exists that someone – a machine learning researcher, the CEO of an industry company, an expert – knows exactly how to manage AI-related risks in all contexts, but they don't work at the organization who needs the answers. The truth is that today, much about the implementation of AI systems is still in the artisan phase - including risk management. Applying new algorithms to real-world problems and real-world datasets is hard and it's challenging to know the risks that will emerge over time. More: https://insights.sei.cmu.edu/blog/5-ways-to-start-growing-an-ai-ready-workforce/	Structuring the framework to foster curiosity and acknowledge the inherent complexity in risk management processes can help to encourage organizations to think broadly about recruitment and workforce diversity. A single person cannot cover all potential risks, and instead organizations should focus on identifying individuals who can reach across different boundaries within a system to track down an answer.
12. The extent to which the Framework should include governance issues, including but not limited to make up of design and development teams, monitoring and evaluation, and grievance and redress.	10	Carnegie Mellon University - Software Engineering Institute	Rachel Dzombak, Ramayya Krishnan, Carol Smith, Brett Tucker, and Nathan VanHoudnos		Donella Meadows, key systems thinking leader said, "Pay attention to what is important, not just what is quantifiable." Governance structures and issues for AI systems must take into account what is important - and certainly the people that create and develop systems as well as system evaluators are critical to the integrity and responsibility of systems. Such teams play a role in mitigating potential risks, challenging assumptions, and are themselves - a likely system vulnerability. The framework should acknowledge governance and guide how continuous governance structures should both be constructed and supported over time.	