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 April 29, 2022:

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**
General Comment	**1**	**SEI - CERT**	**Dr. Grant Deffenbaugh**	**General**	**What happens if the AI were to start acting in its own self-interest? At what point does the AI become intelligent enough to gain "rights"? How does all of this effect how risks are managed?**	**We would like to get NIST's thoughts on this as well as partner to find answers to these challenging questions.**
General Comment	**2**	**SEI - CERT**	**Dr. Grant Deffenbaugh**	**6.1**	**We believe that it would be beneficial for NIST to map accountability/culpability in this step and not leave it solely to the governance in section 6.4.**	**Map accountability/culpability in this step and not leave it solely to the governance in section 6.4.**
General Comment	**3**	**SEI - CERT**	**Dr. Grant Deffenbaugh**	**General**	**It becomes difficult to assign blame with AI's. Since the AI's are learning who is at fault? The designer for not foreseeing a problem. The operator for not training the AI, or maintaining it correctly? The user for having the AI do something outside of its specifications? Perhaps even the AI itself depending on its level of intelligence?**	**We would like to get NIST's thoughts on this as well as partner to find answers to these challenging questions.**
General Comment	**4**	**SEI - CERT**	**Brett Tucker**	**Section 1, Text Box at Line 21**	**We understand that this is a "Risk Management Framework" for artificial intelligence. The NIST RMF is a great framework in use by many. Unfortunately, your title may lead a lay person to misinterpret and believe that this standard is a direct update or closer association. The text box, as it reads specifies that this AI RMF is exactly NOT following the NIST RMF at all. This tends to cause confusion. This does not dismiss the statement made in Section 2 text box line 15, item number 5 of page 3. We greatly applaud the ability of this framework to mesh with others. The question comes down to when is it required where the NIST RMF is required.**	**Consider modifying the title of the framework or document to be "A Framework for Managing AI Risk" or other title where the notion of the NIST RMF is not confused.**
Risk Framing	**5**	**SEI - CERT**	**Brett Tucker**	**Section 4, Lines 9 - 17**	**We applaud and greatly respect the societal risks that must be addressed with this framework. However, we would also like to see adverse impacts related to general operations and the overall resilience of organizations. If AI is to be used by the public as much as the private sectors in operational environments, we believe that there should be significant emphasis on building trust in AI such that operational disruptions (intentional or otherwise) must be mitigated to bridge the gap of trust for use of this technology. This suggestion will provide greater tie-in to the taxonomy seen in Figure 3 of Section 5 under "Technical" risk characteristics. Could this be a greater call to incorporate the technical aspects into the risk framing as well?**	**Consider and add discussion to the framing of the risk that accounts for operational resilience with the implementation of AI technologies.**
Risk Measurement	**6**	**SEI - CERT**	**Brett Tucker**	**Section 4.2, Lines 7-18, Section 6.2**	**We agree with this section. However, there is one other element that NIST may want to consider in terms of qualitative and quantitative risk measurement--secondary risk impacts. Initial business impacts may be more apparent in some situations, and as stated, this may not even be true. The issue is compounded when considering secondary impacts. Examples may include damage to reputation.**	**Consider mentioning the additional challenge of accounting for secondary risk impacts in section 4.2.1.**

Risk Thresholds	**7**	**SEI - CERT**	**Brett Tucker**	**Section 4.2.2, Lines 7-21**	**We agree with this section. There is additional opportunity here for NIST to remind and instruct the risk community to continually review, analyze, and update their organizational risk appetite in accordance with the shifts in technology and policy. More specifically, this may be a good point to advise organizations to establish "tripwires" or "indicators" that invite these reviews. For example, if a new development in AI application comes to light, organizations should be reminded to review their current risk appetite statements to determine if they are applicable in the new context. This move will also be a good reflection of the signficant elements found in the "Plan" step of the NIST RMF.**	**Consider updating section 4.2.2 to call for more regular review of organizational risk appetites as AI technology and its applications evolve.**
Organizational Integration	**8**	**SEI - CERT**	**Brett Tucker**	**Section 4.2.3, Lines 32-33**	**We noted the message delivered in stating that "Small to medium-sized organizations face different challenges…". However, the statement ends there. We would greatly appreciate additional thoughts here. What are the additional challenges for smaller organizations not necessarily experienced by larger organizations? Do we have data or anecdotal evidence that demonstrates this?**	**Please provide additional thoughts on how small and medium-sized organizations may have different challenges in AI RMF implementation. Could there be a difference in resources--people, expertise, money, etc.? Could there be differences in application?**
Manage and Govern	**9**	**SEI - CERT**	**Brett Tucker**	**Section 6.3 and 6.4**	**We recognize the significant importance of supply chain risk in this framework, as most organizations will seek out and procure AI related technologies. It may be worth tying together or recognizing the overlap of the "Manage" and "Govern" process areas through Supply Chain Risk(s). For example, "external stakeholders" are called out in the table for section 6.2. This is a great point where a text box may call out this overlap.**	**Suggest making stronger connections in the overlap of the "Manage" and "Govern" activities via supply chain risk management principles.**
General Comment	**10**	**SEI - CERT**	**Dr. Shing-hon Lau**	**Section 4**	**Is there room for discussion about the risk of externalities caused by the deployment of an AI system? One can imagine a scenario where use, especially widespread use, of an AI system may disadvantage those who are either unable or unwilling to interact with the AI system. For example, the AI may only be available using a webpage, perhaps negatively affecting those without access to a personal computer.**	**Consider adding an explicit discussion of potential externalities caused by deployment of an AI system. There may also be opportunity to discuss whether the capability to interact with an AI might be related to concepts of fairness.**
General Comment	11	SEI - CERT	Dr. Shing-hon Lau	Section 5.1	The discussion provided in subsections 5.1.1 - 5.1.4 focuses heavily on the technical characteristics of ML models. However, ML models are virtually never used in isolation in any real application. It is far more common to see an Ai system consisting of entire pipeline, constructed for the purpose of advancing some organizational aim. In its most basic form, this pipeline runs from data collection at an initial set of sensors and through data pre-processing stages before arriving at a ML model. The output of this model is then post-processed in a decision-making or action stage. Each of these other stages may themselves contain ML models or will contain "dumb" rules that will interact with the primary ML model. The objective should be to assess the risk of the entire AI system, not just the underlying ML model. Accuracy might be achieved by carefully controlling incoming data, rather than by improving a model. Reliability and robustness might be achieved by filtering out "bad" decisions from a ML model with hard-coded rules. Resilience may be achieved by validating inputs and outputs, rather than by using a resilience model.	I would suggest edits to clarify that it is the AI system that should be tested, and not just the ML model. If desired, a distinction can be drawn between the manner in which the technical characteristics of a system might be evaluated, as compared to the manner in which the technical characteristics of a model might be evaluated.
General Comment	12	SEI - CERT	Dr. Shing-hon Lau	Section 5	AI systems are often employed in contexts where they are expected to learn over time to accommodate the particulars of the environment where they are deployed. There does not appear to be any discussion in this section about how to evaluate risk over time as the AI learns.	I would recommend a subsection dedicated to the discussion of how AI systems (and underlying ML models) may drift over time as they learn and how risk evaluations may need to be conducted to accommodate that drift.

General Comment	13	SEI - CERT	Dr. Shing-hon Lau	Section 5.2	Usability (or understandability) is one socio-technical characteristic is hinted at throughout this section but is not directly named. The risks associated with AI system are closely tied to whether a human operator can actually use the system in furtherance of an organizational aim. This may be aided by explainability or interpretability in underlying ML models, but an AI system that poorly presents this information to a human operator (say due to poor UIUX) is likely to create more risk than a system that presents this information well.	I suggest adding another subsection describing usability as a critical socio-technical characteristic.
General Comment	14	SEI - AI	Carrie Gardner	Section 1, Section 5, & Section 5.1.4	The terms "security" and "resilience" are used with varying levels of scope and framing. The initial reference in Section 1 (page 1, line 15) refers to "security (resilience)", another reference in Section 5 (page 8, line 7 graphic) refers to "Resilience or Security", and another reference in Section 5.1.4 (page 10, line 12) refers to "Resilience or ML Security". These references should be revised for consistency and clarity. We have two additional considerations on this framing: 1. We are wondering if the framing, "ML Security", is too narrow. AI system security includes security considerations beyond just ML Security concerns. 2. We are wondering if "resilience" should be included in this frame. Resilience as we understand refers to the attribute or property to withstand operational disruption from risk. Resilience in this regard generally refers to an attribute that is risk-neutral - implying that resilience should be a technical design characteristic across multiple existing characteristics.	3 suggestions on this comment: 1. Make sure all references to "security" are consistent. 2. Remove the "ML" adjective when describing security, or expand the discussion of ML security and why this was scoped to that. 3. Remove references to "resilience" (and add "resilience" in as a separate technical design characteristic)
General Comment	15	SEI - AI	Carol Smith	Section 1	Suggest change to "For the purposes of the NIST AI RMF the term artif icial intelligence refers to algorithmic processes that learn from data in an automated or semi automated manner."	Software-driven systems, that can recognize patterns, create predictions, make decisions, and/or generate new content, without being explicitly programmed to do so. Examples broadly include - but are not limited to - automated decision making systems, recommender systems, computer vision systems, natural language understanding and generation systems, and software behind autonomous physical systems (e.g., autonomous cars, robotics). This definition excludes general AI (GAI) or the development of sentient technology, neither of which are possible with current AI capabilities.
General Comment	16	SEI - AI	Carrie Gardner	Section 5.3, Section 5.3.3	Model "transparency" is one approach to achieving trustworthy explainations and reasonable inferences from AI. The current framing in Section 5.3 (lines 1-3) implies that models must be transparent to allow for inspection - this is not the case as presented in Sandra Wachter et al's paper, Counterfactual Explanations Without Opening the Black Box. Model transparency may not also be an achievable goal in situations where intellectual property rights are applied.	Shift framing of Section 5.3.3 to the principal that AI should be "documented". Models should be documented, data should be documented, AI systems should be documented.
