



26 April 2022

National Institute of Standards and Technology (NIST)
U.S. Department of Commerce
100 Bureau Drive Gaithersburg, MD 20899
Via Email: AIframework@nist.gov

Re: Response to the AI Risk Management Framework: Initial Draft

IEEE-USA is pleased to submit comments on the initial draft of the AI Risk Management Framework. The information provided below is a combined effort of IEEE-USA, which represents the approximately 150,000 IEEE members in the United States, and the IEEE Standards Association (IEEE-SA), the leading developer of global technical standards. Our members include engineers and scientists who are actively conducting research and development into artificial intelligence (AI), software engineering, cybersecurity, and advanced computing. IEEE-SA is developing technical standards and frameworks that enable professionals to prioritize ethical considerations in the design, development, and deployment of AI and autonomous systems (hereinafter referred to collectively as AI systems).¹

As a community of users, researchers, developers, and individuals impacted by AI systems, IEEE strongly supports NIST's initiative to guide organizations in the assessment of their risks. Upon its final release in 2023, the AI risk management framework (RMF) will be an influential reference document for several reasons. First, as a public good, entities of all sizes and sectors that lack AI guidelines or procedures can either adopt or adapt this text to their specific needs. Second, as an effort that welcomes input from a multi-stakeholder group of global experts, the AI RMF will set a precedent for best practices on this issue. As such, consumers, governments, and industry will view their adoption as a sign of an entity willing to engage in seriously evaluating the direct and indirect impact of their AI systems. Lastly, this technology's continuous evolution requires a flexible tool able to adapt to its increasing capabilities, uses, and effects on society.

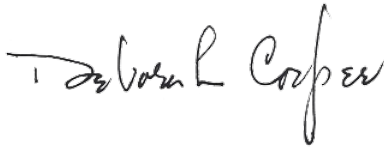
IEEE-USA's input is included in Attachment A and IEEE-SA's input is included in Attachment B.

IEEE thanks NIST for considering these comments as the agency develops a framework to better

¹ See, e.g., The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE, 2019. <https://standards.ieee.org/content/ieeestandards/en/industry-connections/ec/Autonomous-systems.html> (IEEE Ethically Aligned Design); IEEE P7000 Series Standards and Projects addressing topics including transparency, data privacy, and algorithmic bias <https://ethicsinaction.ieee.org/p7000/>; IEEE Model Process for Addressing Ethical Concerns During System Design, IEEE Standard IEEE 7000-2021; IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being, IEEE Standard 7010-2020.

manage risks to individuals, organizations, and society associated with AI systems. We would welcome any further discussions with the agency on these matters. If you have questions, please do not hesitate to contact Erica Wissolik at (202) 530-8347 or e.wissolik@ieee.org.

Sincerely,

A handwritten signature in black ink that reads "Deborah Cooper". The signature is written in a cursive style with a large, stylized 'D' and 'C'.

Deborah Cooper
IEEE-USA President

ATTACHMENT A

IEEE-USA provides the following comments on the first draft of the NIST AI Risk Management Framework

- 1. Implementation Tiers** - IEEE-USA believes that Implementation Tiers are appropriate and necessary tools for mitigating the risks of AI systems and should be included in the final framework. We note that Implementation Tiers were mentioned in the Concept Paper but not in the first draft. Considering the language in the Concept Paper, we are unsure if Implementation Tiers would address our concerns about applying higher levels of risk management techniques to systems that will result in substantial levels of impact on human life and well-being. Therefore, we encourage NIST to both clarify the purpose of the Implementation Tiers and address the need for different levels of risk management given the impact of AI systems.

There are many ways that implementation tiers could be specified. We suggest applying different levels of risk management techniques to systems in relation to their impact on a broad range of stakeholders (e.g., highest level of risk management for systems that impact human life, well-being, and liberty, another level for systems impacting substantial opportunity, etc.). To help enable the application of tiers, we recommend including standards such as IEEE 1012², which specifies different integrity levels for situations where more intensive verification and validation activity, including independent verification and validation, is necessary to protect the public.

- 2. Documenting Risks** - NIST's AI RMF is designed to guide entities assessing the impact of their AI systems. As such, it does not advocate or suggest a particular method for balancing the negative and positive impact of this technology. Each entity is charged with identifying and making a subjective calculus as to which risks are worthy of attention/mitigation or neglect. For categories 3 and 4 of the Map function, IEEE believes that NIST should require entities to document this process with the objective of keeping a record of how risks are evaluated and managed.

In addition, NIST would be well-served by suggesting how organizations should decide whether to deploy a system. This may potentially include developing red lines for technologies that could cause high-impact and long-term effects on society.

- 3. Conflicts of Interest** - AI system end-users would benefit from greater information regarding the alignment of objectives and goals between the technologies that supplement or replace their decision-making. As these systems become more capable, users may rely on them without any awareness of how their outputs could be following objectives that run contrary to the tasks they want to accomplish. The AI RMF would benefit from including guidance under the list of socio-technical characteristics or principles that make transparent whether an AI system is aligned with the objectives or goals of an end user or a third-party.
- 4. General Purpose Systems** - Systems capable of performing multiple tasks are slowly, but surely, emerging in the AI governance conversation. Known in some circles as foundation models, they

² IEEE 1012 <https://standards.ieee.org/ieee/1012/7324/>

represent an important element in the analysis of risks because of their ability to engage with unpredictable inputs. Because of this, the AI RMF would be well-served to include guidance regarding the bespoke evaluation of these systems due to their differences in capabilities as compared to narrow systems.

- 5. Identified Stakeholder Groups (Page 4 -Figure 1)** - Developers of the system fall in the inner circle of the taxonomy, while those purchasing/operating the system are in the next circle. Individuals who are the subject of decision-making by AI systems are placed only in the outer circle (General Public). We suggest modifying this taxonomy so that individuals who are directly impacted by a system are not consolidated with the general public who may not experience the impact of an AI system.

ATTACHMENT B

IEEE Standards Association provides the following comments on the first draft of the NIST AI Risk Management Framework

Comment
<p>Regarding Consultation Question 1 (Q1) and (Question 7) Q7 the AI RMF is a good start, and we commend the AI RMF in terms of how it is positioned, but it may not adequately cover and address AI risks with the appropriate level or degree of specificity as stands. With guidance and other supporting documentation proposed by the framework, the AI RMF will likely become a richer resource but there are a few points of note that are worth highlighting as they are to a degree still missing and could make the AI RMF become more fulsome.</p> <ol style="list-style-type: none"><li data-bbox="228 764 1373 1136">(1) It is not safe to say that risks which are pertinent to any software or information-based system be dealt with solely by separate guidance. While it is important not to reinvent the wheel where existing guidance already applies or is applicable, it is vital to note that the very nature of AI and the context of an AI's application can alter the positioning of the external guidance related to areas such as cybersecurity, privacy, safety, and infrastructure. The AI RMF needs to encompass and adapt and adopt the other guidance specifically in the context of AI. Without doing that organizations will be left to interpret and translate the external guidance for themselves and potentially missing some key AI risks that overlap and converge.<li data-bbox="228 1142 1357 1346">(2) The definition of risk is too narrow in that it needs to map (a) existing risks, (b) opportunity cost risks, (c) consider at pre-design stage whether the AI itself by being allowed to be developed / deployed itself poses a risk (redlining), and (d) risk includes not just the likelihood of a given impact but potential impact and unintended impact and needs to account for the severity of the impact.<li data-bbox="228 1352 1268 1430">(3) Environmental risks posed by AI have not been tackled here. Environmental sustainability ought to be a tenet of the risk management framework.<li data-bbox="228 1436 1365 1724">(4) The AI lifecycle as currently portrayed needs to recognize formally the stage of pre-design and of "decommissioning" which in and of itself poses AI risks. An example of a decommissioning AI risk would be the dependency created on the AI by internal and external AI systems / stakeholders as well as society. Another example would be the duration over which decommissioning is managed due to the need to meet regulatory accountability and auditability needs. Decommissioning ought not just be seen as a risk management technique alone.<li data-bbox="228 1730 1349 1892">(5) While we await further guidance and supporting documents to follow, it does seem like techniques and tools to help organizations measure AI risk once it is identified is lacking. Measurability and having commensurate metrics to understand the potential impact of AI risk (socio-technical risk) can pose real stumbling blocks to organizational

roll out and/or consistent application of a risk management framework.

(6) The Govern section could benefit from adding categories concerning internal whistleblowing of AI risks and external redress mechanisms associated from harm caused by AI risks, and greater emphasis on an organization with a culture that welcomes feedback constructively and does not seek to hide or avert criticism.

(7) Overall, the framework is currently very high level and could do with more depth and illustration through use cases.

Q2. The flexible and voluntary nature of the AI RMF is made clear, but how and when it is to evolve is not currently clear. We suggest that it include a clear statement shaping periodic reviews and by whom, and/or whether intervening case law, use cases, AI risk incidents might also trigger an update to the AI RMF. Those who participate in applying the AI RMF may need to have a mechanism to be kept up to date with or alerted when changes are made to the AI RMF as it evolves so that those dependent upon its guidance and best practice can adjust and modify their internal practices in line with it.

Q3. We suggest that more could be done to enable organizations to be in a better position to make decisions following the AI RMF guidance. This may not be something the framework guidance itself can do. This will likely be more about NIST's role in contributing to the wider AI assurance ecosystem. How will NIST decide to help organizations build their own competence, how will NIST help educate next generation AI risk management resources to enable organizations to build capability? We anticipate that organizations are also going to need to see the RMF as a meaningful exercise for them in which to expend their resources and capacity.

Q4. The functions, categories, and subcategories are clear but are not comprehensive. As it currently stands these are only examples. Over time we would expect to see the AI RMF build up a repository of AI risk functions, categories and subcategories aligned with relevant example use cases. NIST might benefit from gleaning more broader functions, categories, and subcategories from internationally published Algorithmic Impact Assessments.

Q6. The structure of the AI RMF is broadly in line with what other jurisdictions and public sector bodies are promoting as good practice.

Q8. A companion document citing risk management practices aligned to sector and Ai context would be useful.

Q9.

(a) Nothing is mentioned in the AI RMF about what an organization ought to do with its findings. There are greater calls for organizations to either provide AI risk ratings as a result of their risk management endeavors or to be required to publish their Algorithmic Impact Assessments as a part of wider transparency reporting.

(b) While reference to the unique position of Start-ups and SMEs was made, nothing about their specific predicament was addressed in the AI RMF. Specific guidance for such organizations would be recommended.

Comments to Specific Sections of the AI Risk Management Framework Draft

Page Number	Line Number	Comment
4	5, 22	Figure 1. Include ‘future generations’ representation in stakeholder considerations.
4	5, 22	Figure 1. Include consideration of ‘direct stakeholders’ and ‘indirect stakeholders.’ HOW TO READ... (washington.edu)
1, 8-12		Include definition and reference to ‘controllability’ and ‘control considerations’ of the system in the Taxonomy, i.e., Control over AI systems. See Annex E and F, IEEE 7000TM-2021 , for example components of the system that might be within or without external control of the risk manager.
1, 8-12		Include definition and reference to ‘observability’ which is related to and necessary for control of the system in the Taxonomy, See Annex E and F, IEEE 7000TM-2021 , for example, it is required to be able to observe ethical risks and issues of the system. (Note implied reference to ‘observability’ on p10, lines 28-30)
1, 8-12		Include definition and reference to ‘technical measurability’ which is related to and necessary for control of the system in the Taxonomy, See Annex E and F, IEEE 7000TM-2021 . (Note implied reference to ‘technical measurability’ on p10, lines 28-30)

1, 6, 8-12		<p>Include definition and reference to both ‘System Elements’ and ‘System of Systems’ which is related to and necessary for control of the system in the Taxonomy, as well as types of systems. See Annex E and F, IEEE 7000TM-2021, as well as Figure G-1 from ISO/IEC/IEEE 15288:2015</p> <p><i>In general, a higher observability of and control over ethical issues in constituent systems, as in a directed SoS, increases the organization’s capability to include consideration of ethical values during system design and other systems and software engineering processes. (From Annex E, IEEE 7000TM-2021)</i></p>
10	18	Socio-Technical Characteristics should include consideration of Legal, Social and Environmental aspects. See Annex D, IEEE 7000TM-2021
10	28-35	‘Unlike technical characteristics... appropriately’. We suggest that this statement be reworded to remove ‘and cannot yet be measured through an automated process’. Recommend review of IEEE 7000TM-2021 . For example, safety characteristic (5.2.4) is highly observable, measurable, and controllable, including by automated processes.
10	34-35	We suggest additional wording ‘to ensure that risks arising in social, <i>environmental and legal</i> contexts are managed appropriately’
11	32	Privacy can be considered a societal value itself. This is different from the Socio-Technical Characteristics defined as a ‘taxonomy’ (p.10). We suggest removing Privacy from section 5.2, and move to Section 5.3, as Privacy is an ethical value.
6, 12	23, 27	We suggest including a definition of ‘ethical values’ or ‘human values’, see IEEE 7000TM-2021 for an explanation.
5, 6	19-22	‘Risk is a measure of the extent to which an entity is negatively influenced...’. Should include ethical values as being at risk, included in ‘individuals, groups, or communities.... Organizations’.
15	18	Table 1. We suggest changing ‘the specific set of users’ to ‘the specific set of stakeholders’.

15	7	Figure 6. We suggest changing ‘Deployment: user feedback’ to ‘Deployment: stakeholder feedback’.
15	18	‘1. System requirements are elicited and understood from relevant stakeholders...’ and ‘4. Risks and harms to individual....’ Please note that system requirements should be derived through the process of elicitation and risk analysis on stakeholders’ ethical values and the context. We suggest an update to reflect the elicitation and risk analysis process described in IEEE 7000TM-2021 .
15	18	We suggest changing ‘potential users’ to ‘potential stakeholders’
17	1	Table 2, #1, we suggest adding additional wording ‘Elicited system requirements are analyzed for <i>technical and ethical value related risks</i> ’.
6	Figure 2	We suggest that “environment” should be included - a risk assessment not taking the impact of energy consumption into considerations ignores a very significant environmental concern.
15	Figure 6	It is during use and at decommissioning that risks, and impacts can occur. Not including a complete life cycle of an AI system for a risk assessment is in danger of missing where risks occur.
18	3	We suggest that the full life cycle of the AI be considered for risk and impact.
7	Figure 1	Insurers may be an important stakeholder for consideration. In addition, we suggest that the environment be considered. We suggest that this is done in a circle outside of the “General Public.”
5	19	The risk is adequately framed as potential negative influence but please note that this may contradict ISO31000 conception of risk which is regarded as an effect of uncertainty that can have bipolarity.

5	26	It is suggested that this framework should address both harm and benefit as potentialities of risk, but we need to be consistent with the earlier definition of risk, i.e., negative outcomes hence the need for adopting correct vocabulary to refer to all potential negative outcomes as risk and all desirable positive outcomes as reward else there is confusion and contradictions in calling event positive or negative risk
5	18, 26	This framework seems to regard risk as potential negative outcomes (page 5, line 18). A universal management framework should cater for hazards/threats as well as opportunities, i.e., Risk & Reward. In this spirit, we suggest NIST review the title of the framework so that it is extended or emphasis on negative impact/outcome considered for removal since the framework is presented as catering for harm minimization and benefit maximization (page 5, line 26)
6	19	Apart from thresholds, there are also tolerability criteria which are less mechanistic than a numerical threshold triggering action.
14	Fig 5	Measurement is evaluation and that is not the same as assessment.
16	3	Qualitative or quantitative evaluation not assessment.
17	Table 3	While it is accurate that Management involves a measure of judgment/assessment, ideally risk assessment is neither within the capacity nor forte of managers.

About IEEE SA

IEEE SA, a globally recognized standards-setting body within IEEE, develops consensus standards through an open process that engages industry and brings together a broad stakeholder community³. IEEE standards set specifications and best practices based on current scientific and technological knowledge. IEEE SA has a portfolio of over 1,900 active standards and standards projects under development.

³ <http://globalpolicy.ieee.org/wp-content/uploads/2016/05/16011.pdf>