ITI Promoting Innovation Worldwide

January 25, 2022

Via email to: AIframework@NIST.gov

RE: ITI Response to National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework Concept Paper

The Information Technology Industry Council (ITI) appreciates the opportunity to continue its engagement with the National Institute of Standards and Technology as it seeks to develop an *Artificial Intelligence Risk Management Framework.* As such, we are pleased to provide comments on the *AI Risk Management Framework Concept Paper.*

ITI represents the world's leading information and communications technology (ICT) companies. We promote innovation worldwide, serving as the ICT industry's premier advocate and thought leader in the United States and around the globe. ITI's membership comprises leading innovative companies from all corners of the technology sector, including hardware, software, digital services, semiconductor, network equipment, and other internet and technology-enabled companies that rely on ICT to evolve their businesses. Artificial Intelligence (AI) is a priority technology area for many of our members, who develop and use AI systems to improve technology, facilitate business, and solve problems big and small.

ITI is actively engaged on AI policy around the world. We issued a set of *Global AI Policy Recommendations* in 2021, aimed at helping governments facilitate an environment that supports AI while simultaneously recognizing that there are challenges that need to be addressed as the uptake of AI grows around the world.[1] We have also actively worked to inform NIST's efforts to foster trust in AI technology, including responding to NIST's RFI on an AI Risk Management Framework.[2]

ITI and our members share the firm belief that building trust in the era of digital transformation is essential and agree that there are important questions that need to be addressed with regard to the responsible development and use of AI technology. As this technology evolves, we take seriously our responsibility as enablers of a world with AI, including seeking solutions to address potential negative externalities and helping to train the workforce of the future. To be sure, our members are aware of and are taking steps to understand, identify and treat the potential for negative outcomes while leveraging

[1] Our complete *Global AI Policy Recommendations* are available here: https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf

[2] See ITI response to RFI on AI RMF here: https://www.itic.org/documents/artificial-intelligence/NISTRFIonAIRMFITICommentsFINAL.pdf

Global Headquarters
700 K Street NW, Suite 600
Washington, D.C. 20001, USA
+1 202-737-8888

Europe Office
Rue de la Loi 227
Brussels - 1040, Belgium
+32 (0)2-321-10-90

info@itic.org
www.itic.org
@iti_techtweets

opportunities that may be associated with the use of AI systems. As such, we appreciate that NIST is working to establish an AI Risk Management Framework (RMF) and that we have the opportunity to provide input on the initial concepts of this framework.

Below, we highlight some recommendations that we believe will be helpful in strengthening the AI RMF. Following that, we provide feedback on the questions NIST poses in the Concept Paper, along with suggested line edits.

Overarching Recommendations

At the outset, we provide several general thoughts for NIST to consider as it seeks to build out the AI RMF.

NIST should seek to maintain coherence with prior works, clearly establishing a linkage between the AI Risk Management Framework and the Cybersecurity and Privacy Frameworks. Both cyber and privacy-related risks need to be considered in the context of managing AI risk more broadly, so it would be helpful for NIST to articulate more clearly what the overlap or interplay between all of these Frameworks looks like and provide a way in which users can understand how all three Frameworks, and the underlying principles that guide them, can be used together. In particular, although risk management and risk treatment approaches may differ, these Frameworks may share common principles (such as context-driven analysis, the importance of leveraging international standards for risk management and treatment).

NIST should seek to leverage and align the RMF with standards that are currently under development in international standards bodies (ISO/IEC JTC 1). For example, ISO/IEC DIS 23894 - Information technology — Artificial intelligence — Risk management is currently under development. In seeking to align with international standards, NIST should consider updating terminology in the RMF to be consistent with that standard. For example, NIST uses the phrase “high-stakes,” but we encourage it to replace “high-stakes” with “high-risk,” while also providing criteria for users to help determine what actually constitutes a high-risk AI application. We explore this further in response to the questions below.

In considering risks, NIST should clarify how risks differ for human facing and non-human facing AI systems, as well as appropriate risk evaluation criteria. The Concept Paper focuses on AI applications that are human facing, which is an important area. However, a large number of AI applications are not human facing (e.g., analysis of weather information, defects on the factory floor, bottlenecks in networks, or state of the roads) and will have different types of risks from human facing systems.

NIST should seek to maintain and foster consistency internationally to the extent possible. As we noted in our response to the RFI on the AI RMF, international consistency is key, particularly as countries around the world are beginning to consider how to address

risks that may stem from the use of AI (or alternatively, how to harness the benefits that AI will bring). Policy and regulatory divergence pose real risks to the socioeconomic benefits and opportunities of data-driven technologies such as AI, where fair, accurate, fit-for-purpose models depend on access to large, diverse data sets that can flow across borders. Taking into account and seeking to align frameworks to the greatest extent possible will help to ensure interoperability and avoid fragmentation with approaches that other localities, states, or countries may be taking to address AI risk management.

NIST should add a function that accounts for contingencies. We note that as currently envisioned, the functions do not seem to account for contingencies, other than a brief reference in the context of the proposed "Govern" function. In cybersecurity, for example, practitioners do their best to avoid, mitigate, share, transfer, and accept risks. However, organizations also establish incident response practices for the inevitability that incidents do occur. In the same way, organizations should also ensure they are adequately prepared to respond should they be unable to avoid, mitigate, transfer, or accept an AI-related risk. We encourage NIST to develop a Respond (or similar) function, similar to the approach taken in the Cybersecurity Framework, which would map to practices that organizations might undertake to respond to an AI-related incident. While we understand that the broader Govern function may be intended to capture activities such as response and contingencies, in the AI context if may be appropriate to include both Respond and Govern functions. Furthermore, it might be useful to create a database with best practices gathered from the results of such a Respond function so that organizations can leverage such data to anticipate new incidents and deploy mechanisms (some of which may be automated, i.e., MLOps) to consistently check for risk factors. This may also help to encourage stakeholder alignment.

Specific Responses to Questions Posed in the Concept Paper

Below, we also offer discrete thoughts on the questions that NIST poses in the Concept Paper.

- Is the approach described in this concept paper generally on the right track for the eventual AI RMF?

We think the approach taken in the Concept Paper is generally on the right track. We appreciate that the Framework seeks to embody the attributes initially laid out in the RFI, including the flexibility that the Framework as currently envisioned will provide.
With that being said, we are somewhat unclear about how the subcategories within the Framework would be implemented by an organization, given there are many AI standards that are still in development. Indeed, it is important to recognize that the community is in a fundamentally different position than it was when the Cybersecurity Framework was developed (e.g., decades of cybersecurity standardization work, which could be mapped to the Categories and Subcategories in the Framework). We continue to believe that it would

be useful for NIST to conduct a mapping exercise similar to that undertaken in NISTIR 8074 Volume 2. The NISTIR identifies a series of "core areas of cybersecurity standardization" and lists relevant SDOs and key application areas, including whether standards were mostly available, somewhat available, or needed across the identified core areas.[3] While NIST has included an outline of this type in the *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools,* one that is more granular would be valuable, so that NIST (and stakeholders) can have greater awareness of where specific standards exist and where they might be needed for core areas of AI risk management.

- Are the scope and audience (users) of the AI RMF described appropriately?

For the most part, we think the scope and audience are described appropriately. However, we offer some specific line edits below.

	Current Text	**Suggested Text**
p. 2, lines 1-3	*A fourth audience who will serve as a key motivating factor in this guidance is (4) people who experience potential harm or inequities affected by areas of risk that are newly introduced or amplified by AI systems.*	*"A fourth audience who will serve as a key motivating factor in this guidance is (4) people or* ***civil organizations who would have expertise and authority to identify and report on*** *potential harm or inequities* ***experienced by individuals when*** *affected by areas of risk that are newly introduced or amplified by AI systems.* ***A fifth audience to consider are regulators, policymakers and other relevant bodies****.* *Rationale:* By adding "civil organizations" and "regulators, policymakers, and other relevant bodies" to the text, this will ensure consistency with proposed and/or forthcoming regulations, ongoing work within ISO/IEC, OECD, and other source documents

[3] *NISTIR 8074 Volume 2: Supplemental Information for the Interagency Report on Strategic U.S. Government Engagement in International Standardization to Achieve U.S. Objectives for Cybersecurity,* available here: https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8074v2.pdf

p. 3, line 6	*"…representative, test and evaluation personnel, end user, and affected communities, depending on the application."*	*"…representative, test and evaluation personnel,* ***personnel responsible for the AI system support post-market to retirement, suppliers, customers, partners and 3rd parties, end user,*** *and affected communities, depending on the.."* *Rationale:* We encourage NIST to add to the list of stakeholders so that it aligns with those laid out in ISO/IEC DIS 23894, Information technology — Artificial intelligence — Risk management. This would again ensure that the Framework is consistent with ongoing international standards work. NIST could also add language to the list that illustrates stakeholders throughout the full AI lifecycle.

- Are the AI risks framed appropriately?

Below we offer specific proposed line edits, as well as general considerations on how NIST frames risk.

Framing Risk

	Current Text	**Comments/Suggested Text**
p. 3, lines 9-10	"Within the context of the AI RMF, "risk" refers to the composite measure of an event's probability of occurring and the consequences of the corresponding events."	*Within the context of the AI RMF, "risk" refers to the composite measure of an event's* ***likelihood*** *of occurring and the consequences of the corresponding events."* *Rationale:* The term "likelihood" is aligned with ISO 31000 definition of risk.[4]

[4] See ISO 31000:2018.

We appreciate that NIST is seeking to incorporate positive outcomes into the definition of "risk." Indeed, it is important to demonstrate that there are many beneficial uses of AI. While we recognize that NIST's definition of "risk" is aligned with NIST SP 800-160 vol. 1, which notes that risk outcomes can be positive (and can therefore also be thought of as an opportunity), we encourage NIST to make clear in conversations with international stakeholders that that is how positive risk should be interpreted. Oftentimes, risk is only associated with the likelihood of a negative outcome. Alternatively, NIST could consider using the word "opportunity" in the Framework itself.

We also note that on p.3 NIST emphasizes utilizing "measurable criteria that indicate AI system trustworthiness in meaningful, actionable, and testable ways." While utilizing measurable criteria is a laudable goal, we think it important to point out that not everything can be measured or might only be able to be described in a qualitative or semi-quantitative manner due to the current lack of measurements or lack of robust and verifiable measurement methods. While we note that on p.5 NIST states that risks should be "analyzed, quantified, or tracked where possible," as in our previous comments on the AI RMF, we emphasize that the AI RMF should explicitly recognize that not all AI risks can be effectively measured. As NIST develops the Framework, it may be helpful for it to categorize measurement criteria that can be used to map specific measurement mechanisms proposed by NIST, as well as individual AI risks and use cases, for each of the categories described: analysis, quantification, tracking, and response.

AI is an emerging technology area, and standards, guidelines, and best practices are still under development. Because of this, we are also still learning about the range of potential risks, their likelihood, and how to measure them. Thus, NIST should also indicate how the RMF might address a situation where such risks cannot properly be "analyzed, quantified, or tracked." We continue to encourage NIST, in developing the AI RMF, to specifically address situations where risk cannot be measured and offer guidance on reasonable steps for treating that risk, without limiting innovation and investments in new, and potentially beneficial, AI technologies. In the same vein, not every measure of risk is meaningful. This may lead to certain harms being overlooked and is thus, also something that NIST should consider.[5]

Furthermore, NIST might consider offering guidance around validation mechanisms for these risk measurement methods themselves. To illustrate this: On p.7, NIST provides examples of risk measurement mechanisms such as confidence intervals – while this mechanism is widely-used to assess the reliability of AI systems, it does not fully address other sources of error including imbalanced data, biases, and cases where prediction intervals may be more relevant. We believe that the qualitative and quantitative metrics NIST recommends for tracking and treating risk should be associated with a consideration for associated controls.

[5] See Fazelpour and Lipton's "Algorithmic Fairness from a Non-Ideal Perspective" (https://arxiv.org/abs/2001.09773).

Finally, in the section on framing risk, NIST seems to leave out consideration of the unique vulnerabilities that AI may be exposed to from adversarial influence. Since these are also risks to developing trustworthy AI, it may be worthwhile to consider the risks of a model using data that has been poisoned during training or the risks associated with deploying a reinforcement learning model in production without monitoring.

- Will the structure – consisting of Core (with functions, categories, and subcategories), Profiles, and Tiers – enable users to appropriately manage AI risks?

We believe the structure is effective. Indeed, this structure has been successfully leveraged by organizations both with the Cybersecurity Framework and the Privacy Framework, so it seems reasonable to replicate the structure here.

However, in further developing the AI RMF, it would be useful for NIST to leverage ISO/IEC 23894 AI Risk Management, in particular: 6.3 Scope, context and criteria, which describes every phase of the Framework specific to AI systems. ISO/IEC 23894 provides detailed tables that breakdown all risk management activities, in accordance with ISO 31000:2018. This will allow organizations to practically integrate the AI RMF in their existing and audited Management Systems and associated RMFs. Alignment is also critical to ensure that terminology is uniform across frameworks, and so that organizations are not trying to adhere to or incorporate multiple frameworks.

We also encourage NIST to consider revising Figure 1, which illustrates the risk management process throughout the AI life cycle, to align with ISO/IEC 5338 – Information technology – Artificial intelligence – AI system life cycle processes and ISO/IEC 23984 Table C.1 Risk Management and AI System Lifecycle. Right now, deployment is construed as one area. However, it might be helpful to further illustrate the phases following deployment, including the post-market phase, which may engender certain risks across a longer period of time, and the retirement phase, which marks the end of the lifecycle and may also have a different set of risks associated with it. NIST might also consider editing Figure 1 to represent AI risk management as a continuous cycle by adding arrows around the perimeter of the insert.

Below, also offer some specific proposed line edits to the text to improve descriptions of the functions and categories.

Functions

	Current Text	**Comments and/or Suggested Text**
p. 4, line 33-34	"…per a qualitative or more formal quantitative analysis of benefits, costs, and risks, and to stop development or to refrain from deployment."	We believe it would be useful for NIST to offer guidance as to what a sufficient qualitative and quantitative analysis entails, as

		such analysis is imperative in managing risks appropriately.
p. 4, line 35-38	**"**NOTE 1: Context refers to the domain and intended use, as well as the scope of the system, which could be 36 associated with a timeframe, a geographical area, social environment, and cultural norms within which the 37 expected benefits or harms exist, specific sets of users along with expectation of users, and any other 38 system or environmental specifications."	"Context refers to the domain and intended use, as well as the scope of the system, which could be associated with a timeframe, a geographical area, **specific users and affected communities or groups, expectations of users and affected groups, pre-existing patterns of advantage or disadvantage between relevant social groups, differences concerning cultural norms and values,** and any other system or environmental specifications." *Rationale:* The definition of context as crafted inadvertently communicates that Framework users can keep their analysis at a high-level, avoiding specificity around existing patterns of harm. Any analysis of context must prioritize a clear understanding of *specific patterns* of harm impacting *specific* groups, not the not the decisionmakers' high-level perceptions of the general social environment.
p. 5, lines 17-21	"Decisions should take account of the context and the actual and perceived consequences to external and internal stakeholders…"	"Decisions should take account the context, along with the actual and perceived consequences to external and internal stakeholders**, including any disparities between those who potentially benefit and those who potentially could be burdened or harmed.** They should consider interactions of the proposed system with the status quo world, address potential stakeholders' burdens in advance of deployment, and make any changes in status quo

		(including other systems, organizational structures, etc.) that may need to be made to ensure benefits are **achieved in an equitable manner,** and risks minimized **and distributed fairly across stakeholders**.**"** *Rationale:* Risks can be minimized because even if consequences for stakeholders are accurately identified, if the beneficiaries of the technology and those who bear the burdens of the technology are two separate groups, and if the beneficiaries have more power and resources than those who are burdened, risk management is more likely to be undertaken in a way that prioritizes the needs of the beneficiaries. Therefore, decisions should take into account not only the context and consequences for different groups but the *relationship* between those groups, including steps that can be taken to overcome any disparities between those two groups.

Categories

	Current Text	**Comments and/or Suggested Text**
p. 7, line 1, example category 2	*"2 AI capabilities, targeted usage, goals, and expected benefits over status quo are understood"*	*"2 AI capabilities, targeted usage, goals, and expected benefits over status quo,* ***anticipated unintended uses*** *are understood"*

		Rationale: These unintended uses are important to understand since they are a potentially significant source of risk and may be challenging to detect or control.
p. 7, line 1, under "Measure"	*"The effectiveness of existing security controls is evaluated"*	*"The effectiveness of existing controls **(e.g., security, environmental, privacy, safety, etc.)** selected to manage identified risks." Rationale:* We think it would be useful to indicate other controls that may be worth evaluating, including those related to safety and privacy, , consistent with the broader consideration of risks in emerging standards such as ISO/IEC DIS 23894.

- Will the proposed functions enable users to appropriately manage AI risks?

On the whole, the functions seem to be sufficiently broad to cover most areas of AI risk. But we believe the functions could better account for contingencies. In cybersecurity, for example, practitioners do their best to avoid, mitigate, share, transfer, and accept risks. However, organizations also establish incident response practices for the inevitability that incidents do occur. In the same way, organizations should also ensure they are adequately prepared to respond should they be unable to avoid, mitigate, transfer, or accept an AI-related risk. We encourage NIST to develop a Respond function, which would map to practices that organizations might undertake to respond to an AI-related incident. In so doing, as we mentioned above, it would be useful to also consider developing a database or other mechanism to log and/or share best practices across organizations, where applicable. This will help to encourage stakeholder alignment. See reference example.[6]

We also think it is important that NIST consider how and where to include guidance on engaging subject domain experts relevant to the model that is being trained or developed. In managing AI risk, it is important to ensure one has domain expertise that can help analyze the model inputs, so that they reflect the domain scope. To mitigate risk, the Framework should encourage including atypical expertise (at least atypical from a software/AI engineering perspective) to be included during design and development to ensure proper treatment of the domain area.

[5] See the Partnership on AI's AI Incident database as one example: https://incidentdatabase.ai/

- What, if anything, is missing?

While the Concept Paper is a good start, there are several areas that would be useful for NIST to further address in further iterations of the RMF. We address some of these areas in our recommendations at the outset, as well as in our responses to the above questions (see, for example, the need to address contingencies).

In our original comments responding to the RFI, we noted that policy prototyping could be a viable approach to co-developing the RMF. We once again encourage NIST to explore the use of policy prototyping, which is an experimentation-based approach for policy development that can provide a safe testing ground to test and learn early in the process how different approaches to the formulation of the AI RMF might play out when implemented in practice, while assessing their impact before the AI RMF's actual release. Policy prototyping involves a variety of stakeholders coming together to co-create voluntary governance frameworks, based on appropriate standards. Developing and testing governance frameworks in a collaborative fashion allows policymakers to see how such frameworks can integrate with other governance tools such as corporate ethical frameworks, voluntary standards, conformance programs such as those for testing and certification, ethical codes of conduct, and best practices. This method has been successfully used in Europe to test an AI Risk Assessment framework, leading to several concrete recommendations for improving self-assessments of AI.[7]

We also note that the Concept Paper does not address risk evaluation criteria. We had previously recommended that NIST develop a methodology that could help stakeholders determine the risk-level of a specific AI use case and then take steps based on that identification to treat that risk. We continue to believe that such criteria will be key to effectively using the Framework. This is something that we have advocated for more broadly, encouraging stakeholders to work together to characterize "high-risk" applications of AI, including by identifying the appropriate roles for AI developers, users, and other stakeholders in making risk determinations. Such a determination is also crucial for helping stakeholders identify specific technological mechanisms that can be used to measure, mitigate, and control high-risk attributes of AI systems, where applicable. We are not saying that NIST should bucket specific uses of AI into a "high-risk" category, but instead that it should develop criteria that can help the relevant roles with responsibilities and authorities to figure out what level of a risk a particular use case may pose. Including illustrative examples may be helpful, with the clear caveat that the examples are just that, illustrative, and not meant as a categorical determination.

Another area that remains absent in the Concept Paper is considerations around privacy risks. At present, there is limited acknowledgement of the privacy risks to users, individuals that may be impacted by the AI system, and those whose data the model is trained on. This

[6] See OpenLoop *AI Impact Assessment: A Policy Prototyping Experiment*: https://openloop.org/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf

is an incredibly important facet of risk management, and one we also commented on in our initial response to the RFI. Organizations face the very real tension between improving AI and protecting privacy and have no real guidance as to how to resolve this tension. It would be useful for the AI RMF to provide a formula for weighing privacy tensions with those of delivering robust and safe AI experiences throughout the AI lifecycle.

We also believe it would be worth more demonstrating the interplay between the Privacy Framework, the Cybersecurity Framework, and the AI Risk Management Framework, in the same way the Privacy Framework illustrates the overlap between cybersecurity and privacy related events. Indeed, both cyber and privacy-related risks need to be considered in the context of managing AI risk more broadly, so it would be helpful for NIST to articulate more clearly what the overlap looks like and provide a way in which users can understand how all three Frameworks can be used together.

NIST might also consider including language in the preamble of the Framework about how the Framework supports USG efforts to advance U.S. leadership in AI and espouses the principles contained in OMB Memo M-21-06, *Guidance for Regulation of AI Applications*, in the same way it illustrates how the Cybersecurity Framework supports USG efforts to improve critical infrastructure cybersecurity in the preamble of that Framework.

While a risk management framework can be useful, we also encourage NIST to state up front that there are certain risks/functions/technical solutions that may be unknown at the time of publication of the initial version of the AI RMF, so that artificial functional boundaries are not inadvertently created. Indeed, we do not want the AI RMF to accidentally stymie innovation.

***

Once again, we appreciate the opportunity to provide feedback to NIST's on the AI RMF. We believe that such a tool will be helpful but should allow for flexibility in updates given the nascent state of technical solutions related to AI. We hope that such a framework can help provide a construct by which to evaluate risk, while also setting forth a methodology for developers and users to utilize in determining the risk associated with a particular use of AI technology. We are equally committed to the responsible development and deployment of AI technology and encourage NIST to view us as a partner. We are always available for additional conversations on this subject.