accenture

April 29, 2022

Dr. Laurie E. Locascio
Director
National Institute of Standards and Technology (NIST)
Re: Artificial Intelligence Risk Management Framework Initial Draft

Dear Director Locascio,

As a leading global professional services company, Accenture provides a broad range of services and solutions in strategy and consulting, technology, interactive, and operations, that span all industries. We combine artificial intelligence (AI) with deep industry and analytics expertise to help our clients embrace these emerging, intelligent technologies confidently and responsibly.

Accenture appreciates the opportunity to provide input on NIST's AI Risk Management Framework (AI RMF) Initial Draft. We commend NIST for its longstanding and ongoing contributions to public private partnerships, notably its Cybersecurity Risk Management Framework. The Cybersecurity Framework is an effective example of public private collaboration to achieve actionable and iterative guidance, standards, and best practices to manage risk and is a model for the AI RMF.

We are encouraged by NIST's work toward developing the AI RMF and have outlined comments and recommendations that we believe will strengthen the structure and approach to risk management outlined in the Initial Draft.

We have actively worked to inform NIST's efforts to develop an AI RMF, including responding to NIST's initial requests for information on the AI RMF[1] and AI RMF Concept Paper, and participating in NIST's March 2022 AI RMF Workshop. We look forward to further participation and collaboration.

Sincerely,

Teresa Tung
Cloud First Chief Technologist
Accenture

[1] Accenture's response to the initial RFI on the AI RMF is linked here: https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0098.pdf

Accenture
Comments to National Institute of Standards and Technology
Artificial Intelligence Risk Management Framework Initial Draft

Introductory Comments

Accenture believes NIST is generally on the right track as it develops the AI RMF and have outlined several areas that would benefit from clarification, amendment, or supplementation in future drafts. We agree with NIST's stated objective of creating and maintaining "actionable guidance" and agree that "cultivating trust and communication about how to understand and manage AI Risks of AI system will create opportunities for innovation and realize the full potential for the technology." We support NIST's approach to risk mitigation, with voluntary risk-based consensus standards and guidance for AI that account for the varying magnitude and nature of consequences. We also applaud NIST for its thoughtful approach to stakeholder comment and inclusion and its understanding that AI can "benefit nearly all aspects of society and our economy."

Accenture recognizes that this Initial Draft is a relatively early step in NIST's deliberative and inclusive process of developing the AI RMF. To ensure NIST remains on the right track, we have outlined several areas that would benefit from clarification, amendment, or supplementation in future drafts and the upcoming companion practice guide. In addition to our answers to the nine specific questions posed by NIST, we have outlined several high-level recommendations that we believe are key to creating an actionable, interoperable, and widely adopted AI RMF that protects communities while fostering innovation:

- **NIST should use the AI RMF to provide actionable, practical guidance:** Leaders, executives, and other stakeholders need actionable guidance on AI risk management that can be implemented across organizations. We believe this is the foremost lens through which the AI RMF should be drafted and considered. The soon to be released companion "practice guide" should be produced with the level of detail necessary to enable organizations to practically implement the AI RMF. We have highlighted several areas of the Initial Draft that would benefit from increased detail throughout this comment letter.

- **NIST should align the AI RMF with other frameworks, standards, and models:** We are encouraged that NIST is writing the AI RMF to be both law- and regulation-agnostic and interoperable with emerging regulation and governance mechanisms. NIST should ensure interoperability between the AI RMF and other existing and emerging regulation and governance mechanisms. Emerging regulation will drive risk management approaches within organizations. To that end, we recommend NIST include practical guidance to help users understand how the AI RMF interoperates with regulation. We believe this inclusion would significantly increase the value of the AI RMF for organizations and other users.

- **NIST should continue to leverage existing best practices:** We commend NIST for leveraging best practices and other successful attributes of the Cybersecurity Framework, which brought a shared understanding of the vernacular and flexible risk management approaches. We appreciate that NIST is broadly following much of the Cybersecurity Framework's structure as it develops the AI RMF but recognize that AI systems' unique attributes could require some amendments to that structure. We also encourage NIST to clearly illustrate the links between the AI RMF and the Cybersecurity Framework and Privacy Framework, noting the strong connection between managing AI-related risks and privacy and cybersecurity risks.

- **NIST should take cues from leading private sector and foreign initiatives:** Given NIST's strong history of public private collaboration, we recommend collaborating closely with Business Roundtable (BRT) and the Information Technology Industry Council (ITI). We also recommend, in certain cases, leveraging best practices from other countries' frameworks. The Monetary Authority of Singapore's (MAS) recently published guidance[2] on AI fairness, ethics, accountability, transparency is an example of the level of detail needed to provide actionable guidance for organizations to manage AI-related risks and operationalization that may be applicable to the NIST AI RMF.

- **NIST should ensure the AI RMF is a flexible, living document:** Like NIST, we believe the AI RMF should be an iterative process, drawing from the success of NIST's Cybersecurity Framework. AI is a rapidly maturing and fast-changing field in which new, transformative technologies are emerging constantly. We support the use of flexible approaches that can continuously adapt to rapid changes and updates in AI applications.

- **NIST should include benchmarking measurements:** We recommend NIST define best practices for performing benchmarking exercises and include recommended thresholds for benchmarking measurements. Concepts like accuracy, reliability, transparency, and bias are never absolute and indeed can be quite subjective.

Specific Responses to Questions Posed in the Initial Draft

1. *Whether the AI RMF appropriately covers and addresses AI risks, including with the right level of specificity for various use cases.*

Accenture agrees with NIST's broad definition of risk as the composite measure of an event's probability of occurring and the consequences of corresponding events, regardless of whether they are negative or positive. This is preferable to other definitions of risk that focus exclusively on negative potential impacts. We also agree that identifying, mitigating, and minimizing risks and potential harms are essential steps toward the acceptance and widespread use of AI technologies.

In our response to the Concept Paper, we recommended that NIST consider the concept of materiality when framing risk in the AI RMF because it is important to consider both the likelihood of an outcome

[2] More information on the Monetary Authority of Singapore's approach to AI governance can be found here: https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

and the context of potential harm. For example, potential risks that are highly likely to happen and highly likely to cause a life-altering or livelihood-altering outcomes are considered high materiality. The financial services industry, in collaboration with a myriad of agencies, has developed this as a scalable way to address risks and communicate them to stakeholders and regulators. We appreciate that in Section 4.1 (page 5, lines 20-21) of the Initial Draft, NIST wrote that risk is typically a "function of 1) the adverse impacts of the circumstance or event that occurs; and 2) the likelihood of occurrence."

Accenture encourages NIST to note that AI-related risks should be compared to existing human processes. We call this the "human-baseline approach" – setting the bar against human legacy systems, not against vague risk without important context. For example, consider an AI solution that can review medical claims and approve benefits with 90% accuracy within hours of filing a claim. Rather than evaluate the risk in the context of the 10% error, benchmark against the current manual process which may only be 75% accurate. While AI systems can certainly be more accurate than human legacy systems, we also note that putting a human in the loop of AI systems can boost accuracy and optimize output. The human-in-the-loop process represents the best and highest use of humans and machines by merging their respective strengths and abilities.

We also recommended in our response to the Concept Paper that NIST consider a step for triage to enable organizations to mitigate risks that are high materiality early in the AI lifecycle and increase the level of specificity for high-risk use cases. AI is increasingly being developed and deployed to critical processes (e.g., healthcare, employment, judicial, policing, etc.) where it could pose risks to safety, privacy, and human rights. Therefore, it is important to evaluate the level of risk posed by AI and its intended application to determine an appropriate course of action for mitigating any existing or potential risks. Organizations can determine the category of risk by asking difficult questions, such as, "what is the size of the potentially impacted audience?" and "how long will the impacts of the use case affect end users?" Organizations need to put in place different processes to deal with low- and high-level risk.

We commend NIST's definitions of the AI RMF's various audiences in Section 3 and offer two additional suggestions to further build out the "Audience" section. First, we recommend that NIST include data owners as part of the "AI system" stakeholder group (Section 3, page 4, line 8), given the tight linkage between data, data governance and AI. Second, we recommend that NIST expand and refine the definition of the "general public" stakeholder group (Section 3, page 5, line 4), by increasing the number of roles listed to include roles beyond consumers, such as citizens, patients, and employees.

2. *Whether the AI RMF is flexible enough to serve as a continuing resource considering evolving technology and standards landscape.*

Accenture believes the AI RMF Initial Draft is flexible enough to serve as a continuing resource considering the evolving AI technology and standards landscape. We strongly encourage NIST to ensure that both the AI RMF and the companion practice guide are flexible, living documents. We believe NIST should strive to emulate the success and continued flexibility of the NIST Cybersecurity Framework. AI, like cybersecurity, is a rapidly maturing and fast changing field in which new, transformative technologies are emerging constantly. We recommend NIST continuously update the AI RMF through public consultation and requests for information (RFI) like its recent RFI on *Evaluating and Improving NIST*

Cybersecurity Resources: The Cybersecurity Framework and Cybersecurity Supply Chain Risk Management.[3]

In general, we support the use of flexible approaches that can continuously adapt to rapid changes and updates in AI applications. We commend NIST for committing to a flexible AI RMF and companion practice guide that evolve with continued stakeholder comment as these fast-moving technologies progress and mature.

With that said, we believe NIST should provide some certainty into this living document by clarifying which specific aspects, (fairness, accuracy, etc.) will continue to be important to NIST as it updates this document. A degree of certainty, in this case regarding the AI RMF's risks, principles, and characteristics, would be useful for organizations and operationalization as AI technologies continue to advance and mature. We believe that it is important to anchor this living, iterative guidance in clear and consistent principles.

3. Whether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks.

The AI RMF, in its current form, is a valuable step toward enabling decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks. We are encouraged that the Initial Draft is strongly aligned with Accenture's current thinking related to responsible AI.[4] However, as it stands, the AI RMF needs a robust companion practice guide to drive adoption and be practically actionable for organizations of all sizes. We strongly believe the AI RMF should be a practically actionable guidance and urge NIST to develop the companion practice guide with this goal in mind.

We believe NIST included the correct principles and characteristics in its taxonomy on page 8: "Trustworthy AI: Risks and Characteristics." However, we encourage NIST to clarify the definitions and boundaries associated with the characteristics and principles outlined in the taxonomy. The current definitions and explanations of the distinctions between technical characteristics, socio-technical characteristics, and guiding principles are unclear. For example, all the topics that NIST defines as "Risks and Characteristics" require human input and could therefore be considered socio-technical characteristics.

NIST could also improve the taxonomy by applying the "Risks and Characteristics" to the AI lifecycle and indicating which topics are relevant when (e.g., in pre-, in-, or post-processing) and which are overarching. Organizing the topics according to lifecycle would address another gap: acknowledging and addressing the trade-offs that decision makers need to consider when prioritizing these topics. NIST should consider exploring the inherent contradictions and subsequent tradeoffs between the characteristics.

[3] NIST's recent RFI on updating its Cybersecurity Framework is linked here: https://csrc.nist.gov/News/2022/rfi-evaluating-and-improving-nist-cyber-resources
[4] More information on Accenture's latest thinking on responsible AI can be found here: https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice

For example, principles such as transparency and security can be contradictory. An overemphasis on the transparency of an AI system could negatively impact data security, particularly if there were requirements to open-up detailed information regarding the training dataset being used and the type of machine learning algorithm, making attacks on machine learning models, such as model extraction, much easier. Additionally, efforts to promote privacy, fairness, and explainability could result in decreased accuracy (e.g., increased error rate).

Generally, the decision-making regarding the risks and characteristics topics are subjective, and as such subjective decisions and their justifications need to be documented for transparency and accountability purposes. One effective way to do so is for an organization to adopt use of datasheets for datasets[5] or model cards for reporting.[6]

4. Whether the functions, categories, and subcategories are complete, appropriate, and clearly stated.

Accenture believes the functions, categories, and subcategories are generally appropriate and clearly stated and will serve as a solid foundation for the AI RMF but are currently incomplete to serve as practically actionable guidance for organizations. We recommend that NIST increase the current level of detail in the next draft, specifically in Section 6: "AI RMF Core." The sheer scope and systemic nature of the work that is required to manage and mitigate AI-related risks requires an increased understanding of AI risks and, in some cases, specific controls. We understand that this is an initial draft and that NIST is working on an accompanying practice guide and a second draft, but we encourage NIST to increase the level of detail for the functions, categories, and subcategories.

Specifically, we recommend NIST explicitly mention that risk thresholds need to be determined, approved, and documented in relevant standards in *Table 4: Example categories and subcategories for the Govern function* (page 19, ID 1). This is one of the more difficult elements of risk management and is a requirement for the three areas laid out in the subcategories for *Govern* in the same table.

More broadly, we recommend NIST look to its own Cybersecurity Framework, which includes specific controls organizations can apply to reduce risk[7] and incorporate similar controls into Section 6 of the AI RMF to help organizations measure and manage risks.

Accenture also encourages NIST to take cues from best practices in other countries' AI frameworks and guidance. Specifically, we encourage NIST to consider MAS' recently-updated guidance on FEAT (fairness, ethics, accountability, transparency) as an example guidance to enable organizations to effectively manage and mitigate risks.[8]

[5] Please see this paper for more information: https://dl.acm.org/doi/fullHtml/10.1145/3458723

[6] Please see this paper for more information: https://dl.acm.org/doi/abs/10.1145/3287560.3287596

[7] Controls in the NIST Cybersecurity Framework are linked here: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#/

[8] More information on the Monetary Authority of Singapore's FEAT principles can be found here: FEAT-Principles-Updated-7-Feb-19.pdf (mas.gov.sg)

MAS established and led an industry coalition, Veritas[9][10], which defined guidelines and operationalization from financial services institutions on the FEAT principles. This includes a recently released guidance in which Accenture helped MAS approach operationalizing FEAT in new ways that may be applicable to the NIST AI RMF:

1. Enabling the scope of protected characteristics to cover local, cultural norms by focusing on personal attributes instead of narrowly defined classes.
2. Providing a normative framework for organizations to articulate their own ethical commitments, with associated measurements for accountability.
3. Aligning the FEAT methodology and checklist questions to AIDA system design, development, and operations – enabling "FEAT by design" to be integrated throughout the product lifecycle.
4. Using a risk-based approach to scale FEAT pragmatically across an organization.[11]

While we recognize that quantity does not in and of itself trump quality, we believe that guidance on managing and mitigating AI-related risks requires a greater level of detail than what was included in the AI RMF Initial Draft. We encourage NIST to increase the level of detail for the functions, categories, and subcategories in subsequent drafts.

5. Whether the AI RMF is in alignment with or leverages other frameworks and standards such as those developed or being developed by IEEE or ISO/IEC SC42.

While jurisdictions around the world, notably the European Union, are pursuing their own regulatory and non-regulatory approaches to AI, NIST should collaborate with the EU and partners within the OECD, to agree to common definitions and terms related to risk. A common lexicon would enable the global AI community to speak the same language, give organizations and society more confidence in AI, and promote greater alignment of standards, frameworks, and models on AI risk management.

NIST should avoid using terms and concepts that conflict with other frameworks, including those published and being developed by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC)[12], as well as the Institute of Electrical and Electronics Engineers (IEEE).[13]

We support the development of voluntary consensus standards that create and safeguard trust at the heart of AI-driven systems and business models while permitting the flexibility for innovation. We encourage

[9] More information on Veritas, the Monetary Authority of Singapore's multi-phased collaborative project with financial industry to bolster internal governance around the application of AI and the management and use of data, is linked here: https://www.mas.gov.sg/schemes-and-initiatives/veritas

[10] More information on the Monetary Authority of Singapore's assessment methodologies are linked here: https://www.mas.gov.sg/-/media/MAS-Media-Library/news/media-releases/2022/Veritas-Document-3B---FEAT-Ethics-and-Accountability-Principles-Assessment-Methodology.pdf and here: https://www.mas.gov.sg/-/media/MAS-Media-Library/news/media-releases/2022/Veritas-Document-3A---FEAT-Fairness-Principles-Assessment-Methodology.pdf

[11] Information on Accenture's work with MAS on the FEAT principles can be accessed here: https://www.accenture.com/sg-en/blogs/southeast-asia-blog/mas-goes-from-principles-to-practice-in-fairness-ethics-accountability-and-transparency-for-financial-services

[12] More information on ISO/IEC JTC1, the joint international standards committee, is linked here: https://jtc1info.org/technology/subcommittees/artificial-intelligence/

[13] IEEE's Artificial Intelligence Systems (AIS) Related States can be found here: https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards.html

NIST to look to the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems[14] and its nine pipeline standards on Ethically Aligned Design as additional examples.

6. Whether the AI RMF is in alignment with existing practices, and broader risk management practices.

Accenture commends NIST for striving to leverage best practices and other successful attributes of the Cybersecurity Framework, which brought a shared understanding of the vernacular and flexible risk management approaches. We encourage NIST to clearly link the AI RMF to its Cybersecurity Framework and Privacy Framework, noting the strong connection between AI-related risks and privacy and cybersecurity risks. As NIST develops the second draft of the AI RMF, we also recommend that NIST assume readers will not have read the Cybersecurity Framework and Privacy Framework and consider including a section (or separate document) illustrating the linkage between the three respective frameworks. We encourage NIST to connect the AI RMF with the Cybersecurity Framework and Privacy Framework to ensure alignment across AI, privacy, and cybersecurity – three areas that often overlap.

We also recommend that NIST take cues from leading private sector initiatives, guidance, and standards in other countries. Given NIST's strong history of public private collaboration, we recommend collaborating closely with the BRT and ITI – and specifically taking cues from BRT's *Roadmap for Responsible AI and Policy Recommendations*[15] and ITI's *Global AI Policy Recommendations.*[16] And, as stated in our response to Question 4, we encourage NIST to consider effective guidance in other countries, including the Monetary Authority of Singapore's Guidance on FEAT (Fairness, Ethics, Accountability, and Transparency) and the UK's AI and data protection risk toolkit.[17]

7. *What might be missing from the AI RMF.*

NIST recently published a robust technical guidance paper focused on bias titled, *Toward a Standard for Identifying and Managing Bias in Artificial Intelligence*, that is related to the AI RMF and greater AI-related NIST efforts.[18] We encourage NIST to look to its *Bias* paper – a practical and detailed guidance – and incorporate its central tenets into the AI RMF, notably the "Recourse Channels" outlined under Section 3.4.1 "Governance Guidance" on Page 43 of the *Bias* paper. The authors suggested organizations

[14] You can access the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems here: https://standards.ieee.org/industry-connections/ec/autonomous-systems/

[15] You can find the BRT's Responsible AI Roadmap and other AI-related policy recommendations here: https://www.businessroundtable.org/business-roundtable-launches-responsible-ai-initiative

[16] ITI's complete Global AI Policy Recommendations are available here: https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf

[17]More information on the UK's AI toolkit is linked here: https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/ai-and-data-protection-risk-toolkit/

[18] NIST's paper, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, can be found here: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

establish the “availability of feedback channels allow system and end users to flag incorrect or potentially harmful results and seek recourse for errors or harms.”

We recommend that NIST include similar guidance for operators to have documented policies and procedures in place for those negatively impacted by by an AI system to seek recourse and potentially receive appropriate and timely redress. While redress is mentioned briefly twice in relation to transparency on page 13 of the AI RMF Initial Draft, we believe it should be emphasized and further developed, potentially under the Initial Draft’s “Govern” function.

Redress and Recourse Channels are especially important for AI systems used in areas including employment and health care where the trajectory of someone’s life could be affected by an AI system. In these areas, there should be higher disclosure, tracking, reporting, recourse, and redress standards. As AI is increasingly being deployed to critical processes where it could pose risks to safety, privacy, and human rights, the importance of redress and recourse channels increases.

Accenture has also identified additional risks that we work with our clients to mitigate and would encourage NIST to consider, including data risk**.** Data risk prompts the need for a fit for purpose assessment that determines whether the data is appropriate to be used (e.g., is it up to date; relevant) and documents of data inclusion or exclusion as well of limitations of the included data that may lead to bias and unfairness when analyzed.

Furthermore, data risk in terms of inclusion/exclusion criteria has implications for fairness. For example, in scenarios when someone did something “wrong” in the past but have paid their dues, made amends, and implemented changes, any of which indicate that fact represented in the data is no longer applicable or a fair representation of that individual, and therefore either should not be included in the dataset to be analyzed or should be minimally weighted so to not have an outsized influence on the output of the analysis. This concept is a form of *reputation bankruptcy*[19] and conveys fairness in the form of forgiveness. Such considerations could bolster NIST’s definition of fairness in the next draft of the AI RMF.

Finally, we encourage NIST to clarify its thinking on meaningful human review of AI technologies and how human review is connected to governance and other obligations for human involvement.

8. *Whether the soon to be published draft companion document citing AI risk management practices is useful as a complementary resource and what practices or standards should be added.*

The soon to be published companion document and NIST AI Resource Center should be a valuable and practically useful complementary resources for stakeholders and users.

In our response to the AI RMF Concept Paper and our introductory recommendations in this comment letter, we emphasized that leaders, executives, and other stakeholders need actionable guidance on AI risk management that can be practically implemented across organizations. We believe this is the foremost

[19] For more information on this concept, please see here: http://blogs.harvard.edu/futureoftheinternet/2010/09/07/reputation-bankruptcy/

lens through which the AI RMF should be drafted and considered. The companion document, if constructed in a way that is practically implementable and released in tandem with a detailed AI RMF, has the potential be an actionable and valuable resource for a wide array of stakeholders. It would distinguish the AI RMF from a plethora of standards and guidance that list high-level principles but ultimately do not enable organizations to translate them into practical and measurable action.

Accenture research found that organizations need to tackle a central challenge when establishing AI governance structures: translating ethical principles into practical, measurable metrics that work for them.[20] Organizations must move beyond defining "Responsible AI" and put those principles into practice.

Specifically, we recommend NIST consider several forms of output including source code, metrics, measures, workshop templates, checklists, case studies and papers discussing their context and use in the companion document.

9. Others?

NIST's recent paper, *Toward a Standard for Identifying and Managing Bias in Artificial Intelligence*,[21] includes some excellent "components" for effective AI system governance. Section 3.4.1 "Governance Guidance," which begins on page 42, builds out these components in detail. Accenture recommends incorporating these components in the AI RMF:

- Recourse Channels
- Policies & Procedures
- Documentation
- Accountability
- Culture & Practice
- Effective Challenge
- Three Lines of Defense
- Risk Mitigation, Risk Tiering and Incentive Structures
- Information Sharing

[20] More information on Accenture's Responsible AI practice and research can be accessed here: https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice
[21] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf