ITI Promoting Innovation Worldwide

September 8, 2021

Mark Przybocki
U.S. National Institute of Standards and Technology
MS 20899
100 Bureau Drive
Gaithersburg, MD 20899

Via email to: AIframework@NIST.gov

RE: ITI Response to National Institute of Standards and Technology (NIST) Request for Information on an Artificial Intelligence Risk Management Framework [Docket Number 210726-0151; NIST-2021-0004]

Dear Mr. Przybocki:

The Information Technology Industry Council (ITI) appreciates the opportunity to submit comments in response to the National Institute of Standards and Technology's Request for Information on its *Artificial Intelligence Risk Management Framework.*

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 and its member companies believe that effective government approaches to AI clear barriers to innovation, provide predictable and sustainable environments for business, protect public safety, and build public trust in the technology.

ITI is actively engaged on AI policy around the world and issued a set of *Global AI Policy Recommendations* earlier this year, 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 engaged with NIST as it has considered various aspects important to fostering trust in the technology, most recently on explainability.

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

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, the tech industry is aware of and is already taking steps to understand, identify and mitigate the potential for negative outcomes that may be associated with the use of AI systems. As such, we appreciate that NIST is considering how to establish an AI Risk Management Framework (RMF) and that we have the opportunity to provide input on what should be included in such a framework to be of most use to stakeholders.

General Thoughts

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

Consider whether the aim is to build a trustworthiness framework or a risk management framework. In some of the questions that are posed in the RFI, it seems that NIST is seeking to build a trustworthiness framework, which in our view would be distinct from, but have some overlap with, a risk management framework. Characteristics that contribute to trustworthiness are one part of risk management but focusing solely on trustworthiness negates other risks that may arise in the AI life cycle. For example, security and privacy are two areas that must be considered in AI risk management but focusing only on trustworthiness may serve to exclude (or only partially consider) these important principles from consideration. Trustworthiness should not be a proxy for risk management. As such, we recommend that NIST clarify how the concept of trustworthiness fits *into* a risk management framework, ideally noting that it is one aspect of risk management.

Consider what "risk" means. We encourage NIST to think about what "risk" means in the context of AI as it develops a risk management framework. There are different sorts of risks that one might associate with an AI system, which may require different mitigations. For example, there are risks to safety (i.e. AI causing an advanced manufacturing system to malfunction, harming someone in the plant) and risks to society more generally (i.e. an AI system that denies a mortgage application based on specific attributes). In seeking to develop a risk management framework for the entirety of AI, NIST also needs to consider that the environments and applications of the technology can differ dramatically, even when the underlying algorithm is identical. Consequently, the magnitude and types of risk/hazard associated with an AI system in one application may be slightly or vastly different than that associated with the same or a similar AI system in another application. Thus, risk consideration must be dynamic and context-sensitive.

Ground the development of the AI RMF in experimentation and evidence through policy prototyping. We encourage NIST to explore the use of policy prototyping as a method through which to co-create and test the AI RMF. Policy prototyping 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 governance frameworks, including regulation and voluntary standards. Developing and testing governance frameworks in a collaborative fashion allows policymakers to see how such frameworks can integrate with other co-regulatory 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.[2]

Specific Responses to Questions Posed in the RFI

Below, we also offer discrete thoughts on many of the questions that NIST poses to stakeholders in the RFI.

1. *The greatest challenges in improving how AI actors manage AI-related risks—where "manage" means identify, assess, prioritize, respond to, or communicate those risks;*

There are a variety of challenges when it comes to managing AI-related risks. At the outset, it is worth noting that because AI is an emerging technology area, standards, guidelines, and best practices are still being developed. When the Cybersecurity Framework was under development, in contrast, the standards, guidelines, and best practices that the Framework mapped to were much more established, having been built upon multiple decades of real-world experience, than they are in the AI space. As such, additional collaborative work needs to be done to develop and mature AI standards and best practices, especially regarding methods for assessing, measuring, and comparing data and AI systems. Keeping this in mind, it may prove to be difficult to build a comprehensive risk management framework populated with fully developed standards at the start– it will likely need to be an iterative process that is updated periodically.

There are a series of other challenges that ITI identified in conjunction with our membership, including:

- *Identifying where bias occurs in the AI lifecycle and assessing how that may impact outcomes.* There is no real consensus around what fairness looks like in the context of AI, nor is there consensus around what reasonable mitigation looks like, which complicates the ability of AI actors to control for this risk.

[2] 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

- *Communicating how an AI system made a decision and determining when providing that explanation might be appropriate.* We previously provided comments to NIST on Draft NISTIR 8312: *Four Principles of Explainable AI,* where we outlined a series of perspectives with regard to explainable AI.[3]
- *Managing risks related to AI security.* AI systems are susceptible to familiar vulnerabilities, like remote code execution and memory code disruption, as well as AI-specific vulnerabilities, such as membership inference attacks and model inversion. Adversarial examples may also negatively impact AI systems.
- *Addressing the tension between improving AI systems and protecting privacy.* As an example, improving state-of-the-art AI currently often means self-supervised and semi-supervised learning. In order to train robust and fair self-supervised models, they need to be trained on massive quantities of inclusive and representative datasets. Often, building these kinds of models is in direct tension with data minimization and purpose limitation privacy principles. We hope that the AI RMF provides a helpful formula for weighing privacy tensions with those of delivering robust and safe AI experiences throughout the AI lifecycle. As such, we ask that the AI RMF provide companies with the ability to document the tradeoffs between state-of-the-art AI and privacy.
- *Acknowledging that many AI harms are still unknown.* As AI is still an emerging technology and applications and use cases are still being explored, AI harms are still largely undetermined, making it challenging to plan for and mitigate risks that may arise from those harms. NIST should consider the risk of stifling innovation and limiting AI applications that may arise from a prescriptive approach that seeks to mitigate risks, but which could be addressed as the technology develops and matures.

2. *How organizations currently define and manage characteristics of AI trustworthiness and whether there are important characteristics which should be considered in the Framework besides: Accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security (resilience), and mitigation of harmful bias, or harmful outcomes from misuse of the AI;*

We believe the characteristics of AI trustworthiness that NIST outlines are fairly comprehensive. However, echoing the point we make at the outset of our paper, we encourage NIST to consider how trustworthiness and related characteristics fit into the risk management conversation. We further request that NIST consider situations where trustworthiness could be a proxy for risk management, and in such instances, where and what elements of trustworthiness overlap with elements of a risk management approach.

3. *How organizations currently define and manage principles of AI trustworthiness and whether there are important principles which should be considered in the Framework besides: Transparency, fairness, and accountability;*

[3] See ITI comments here: https://www.itic.org/policy/ITICommentsNISTIR8312ExplainableAI.pdf

We agree that the principles NIST has outlined are all are important to building a trustworthy AI system. However, we believe that several additional principles should be integrated into NIST's thinking around the AI RMF as well. Indeed, these principles should go beyond trustworthiness to include cybersecurity, privacy, and inclusiveness. These are all foundational concepts that should be considered in the AI RMF.

Cybersecurity and privacy are foundational to trustworthy AI systems and there are multiple ways in which cybersecurity, privacy, and AI interact. When thinking about cybersecurity, NIST should consider not only that AI can be leveraged to enhance cybersecurity, but also that there may be security risks resulting from AI systems that need to be managed. In the context of privacy, transparency -- whereby the providers of an AI solution are able to declare how data is being used -- is important. We discuss this challenge more in response to Q1.

Inclusiveness should also be included as a principle in the AI RMF. Indeed, AI systems should be trained with inclusivity in mind, as this is one way that risks like bias can be managed. Ensuring that datasets are representative of a wide variety of attributes and that inclusivity is considered at every stage of the design and development process is an important outcome that stakeholders should strive to achieve.

5. *Standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles to identify, assess, prioritize, mitigate, or communicate AI risk and whether any currently meet the minimum attributes described above;*

It is worth keeping track of ISO/IEC 23894.2, which is currently under development in ISO/IEC JTC 1 SC 42 and is specifically geared toward AI risk management. ISO 31000 on Risk Management, though not specific to AI, also includes elements that are relevant to this effort.

The AI RMF should explicitly recognize that not all AI risks can be effectively measured and we should not prohibit AI innovation as a result of this fact. 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. The AI RMF should specifically address situations where risk cannot be measured and offer guidance on reasonable steps for mitigating that risk, without limiting innovation and investments in new, and potentially beneficial, AI technologies.

6. *How current regulatory or regulatory reporting requirements (e.g., local, state, national, international) relate to the use of AI standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles;*

Likely the most important potential regulatory regime that NIST should keep in mind as it is developing the AI RMF is the European AI regulatory proposal, which lays out a comprehensive regulatory framework for AI and which implicates standards in its chapters

on conformity assessment. Indeed, the EU proposal lays out a template for the application of the New Legislative Framework – the EU's three legislative acts governing standardization, conformity assessment, and accreditation across industrial goods sectors – to high-risk applications of AI. There are specific provisions which seem to indicate an exclusive reliance on harmonized European standards as a means to demonstrate compliance, and further provisions which grant the Commission powers to adopt common specifications via implementing acts in cases where relevant harmonized European standards do not exist or are found to be insufficient for the protection of "fundamental rights." ITI filed comments on the EU AI proposal, in which we strongly encourage the EU, in establishing a regulatory regime, to rely on voluntary, industry-driven, consensus-based international standards as a means to establish consensus around technical aspects, management, and governance of AI technology, as well as to frame concepts and recommended practices to underpin trustworthiness of AI inclusive of privacy, cybersecurity, safety, reliability, and interoperability.[4] However, should the EU disregard such a recommendation and pursue the adoption for common specifications, NIST should consider the impact on the direction of the AI RMF.

It may also be worth noting regulations released in Shenzhen, China, which may shed light on the direction China as a country chooses to take in governing AI. In June 2021, the city submitted the *Regulations on the Promotion of Artificial Intelligence Industry of Shenzhen Special Economic Zone* to the People's Congress for review. The regulations include provisions around the ethical use of the technology and establish a framework to administer an approval process for AI products and services.[5]

It is also important to note existing regulatory frameworks in areas other than AI, like the EU's GDPR and other global privacy regimes, when considering an AI risk management framework. As there is significant interplay between AI and privacy, NIST should take into account how existing regulatory regimes may impact the management of AI risk. Another area that is likely useful to consider is existing product liability regimes. In the EU, for example, the European Commission is considering if/how to revise product liability rules to account for emerging technologies like AI. While we do not believe NIST should necessarily address issues of liability in the AI RMF, it is nonetheless important to think about how existing product liability regimes may function for AI systems.

7. *AI risk management standards, frameworks, models, methodologies, tools, guidelines and best practices, principles, and practices which NIST should consider to ensure that the AI RMF aligns with and supports other efforts;*

[4] *ITI Views on the European Commission's Artificial Intelligence Act Proposal,* available here: https://www.itic.org/documents/europe/20210806ITIResponsetoEUAIActProposal%5B16%5D.pdf

[5] See news release here: http://jzw.sz.gov.cn/jczxdt/szyw/content/post_702250.html and brief analysis here: https://www.china-briefing.com/news/artificial-intelligence-china-shenzhen-first-local-ai-regulations-key-areas-coverage/

There are several efforts, which we highlight below, that NIST should keep in mind as it seeks to develop the AI RMF. Several of these frameworks emphasize ethics but are important as the ethical design and development of AI plays a role in risk mitigation.

U.S./Domestic

DOD Ethics Principles for AI[6]

The DOD has established a set of ethics principles to guide the combat and non-combat functions of the U.S. military in maintaining its legal, ethical, and policy commitments in the field of AI. These principles include responsibility, equitability, traceability, reliability, and governability.

Principles of AI Ethics for Intelligence Community[7]

The Intelligence Community has also adopted AI Ethics Principles to guide personnel on whether and how to use AI, including machine learning, to further the mission of the Intelligence Community. The principles include respecting the law and acting with integrity, ensuring that AI is transparent and accountable, objective and equitable, human-centric, secure and resilient, and informed by science and technology. The IC has also developed a complementary AI Ethics Framework to guide personnel who are figuring out how to procure, develop, use, design, or consume AI.[8]

Office of Management and Budget AI Regulatory Guidance[9]

We urge NIST to also keep in mind the broader OMB *Guidance for the Regulation of AI Applications*. Although NIST is not intending to develop regulation, the principles contained in the OMB memo provide a useful backdrop to frame broader federal AI efforts, including those being undertaken by NIST. In particular, principles related to risk assessment and management, flexible approaches to AI risk management, and public trust in AI are all relevant to the effort to establish the AI RMF.

Global

Countries and organizations around the world have also developed ethics and other frameworks to help guide the development and use of AI. NIST should seek to take these into account to facilitate alignment and interoperability of various AI frameworks to the extent possible. The OECD AI Observatory is a repository that references various national and global efforts and is worth referencing as NIST develops the AI RMF.

---

[6] *Ethical Principles for Artificial Intelligence,* available here: https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

[7] *Principles of Artificial Intelligence Ethics for the Intelligence Community*, available here: https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf

[8] *Artificial Intelligence Ethics Framework for the Intelligence Community*, available here: https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf

[9] *Guidance for the Regulation of Artificial Intelligence Applications,* available here: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

JTC 1 SC 42 Standards Work

We agree with NIST's suggestion that the AI RMF should be consistent with other approaches to managing AI risk and take into account existing standards, guidelines, and best practices. We specifically note the work of ISO/IEC JTC 1 SC 42 in our *Global AI Policy Recommendations*, including the work it is doing on the Artificial Intelligence Management System (AIMS) standard, which will cover processes with the development or use of AI, including related to bias, fairness, inclusiveness, safety, security, privacy, accountability, and explainability, all characteristics that NIST references in its RFI. Leveraging this standard and others that are currently under development in SC 42 around terminology, reference architecture, governance of AI, and trustworthiness will help to improve interoperability and facilitate alignment of approaches to managing AI risks.

IEEE Position Statement on AI[10]

The IEEE announced a position paper that urges governments to adopt policies that increase AI technical expertise within governments and foster greater government access to academic and private sector technical expertise; support R&D; ensure public welfare and provide an effective legal and regulatory framework for AI development, application, use, and monitoring; and facilitate public understanding of and discourse around AI. The position statement also points to the P7000 series of IEEE standards that are being developed and are worth taking into account.

IEEE 7010-2020: Assessing the Impact of AI on Human Well-Being[11]

The IEEE 7010-2020 has also developed a standard that sets forth a comprehensive conceptual framework addressing universal human values, data agency and technical dependability with a set of principles to guide Autonomous and Intelligent Systems creators and users through a comprehensive set of recommendations.

European Commission High-Level Experts Group Ethics Guidelines for Trustworthy AI & AI Assessment List for Trustworthy AI[12]

The European Commission's High-Level Experts Group (HLEG) established a set of seven requirements than an AI system must meet in order to be considered trustworthy, enshrined in the *Ethics Guidelines for Trustworthy AI*. These requirements include: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability.

---

[10] *IEEE Position Statement on Artificial Intelligence,* available here: https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf

[11] *IEEE 7010-2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*, available here: https://standards.ieee.org/standard/7010-2020.html

[12] *Ethical Guidelines for Trustworthy AI*, available here: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai; *Assessment List for Trustworthy Artificial Intelligence,* available here: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

The group has developed a complementary *Assessment List for Trustworthy AI*, which is a self-assessment tool that developers can use to translate the *Ethical Guidelines* into an actionable checklist.

Considerati Artificial Intelligence Impact Assessment (AIIA)[13]

The AIIA is a tool that can help organizations find the right framework of standards and aid in deciding the relevant trade-offs. The AIIA offer concrete steps to help organizations understand the relevant legal and ethical standards and considerations when making decisions on the use of AI applications.

Australia AI Ethics Framework[14] & AI Action Plan[15]

Australia developed an *AI Ethics Framework* to guide businesses and governments to responsibly design, develop and implement AI. The framework consists of eight voluntary principles to ensure AI is safe, secure and reliable. It encourages a focus on human, societal, and environmental well-being, and states that AI systems should maintain fairness, uphold privacy and security, take a human-centric approach, be safe and reliable, ensure transparency and accountability, allow for contestation of a decision, and be accountable.

Australia is also in the process of implementing its *AI Action Plan*, which is a comprehensive, strategic approach to responsibly developing AI in Australia. It focuses on four areas: developing and adopting AI to transform Australian businesses, creating an environment to attract the world's best AI talent, using cutting edge AI technology to address pressing challenges in Australia, and making Australia a global leader in responsible and inclusive AI. Both of these efforts are worth taking into account as NIST seeks to develop the AI RMF.

Singapore Model AI Governance Framework[16]

Singapore has also developed an AI Ethics Framework, which translates existing AI ethical principles -- accountability, accuracy, auditability, explainability, fairness, human centricity and well-being, human rights alignment, inclusivity, and progressiveness -- into recommendations that organizations can adopt to ensure responsible deployment of AI systems. The framework primarily focuses on internal governance structure and measures, human involvement in AI-augmented decision-making, operations management, and stakeholder engagement.

---

[13] *The Artificial Intelligence Impact Assessment*, available here: https://www.considerati.com/static/default/files/documents/pdf/Artificial%20Intelligence%20Impact%20Assessment%20-%20English[2].pdf

[14] *Australia's Artificial Intelligence Ethics Framework,* available here: https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework

[15] *Australia's AI Action Plan*, available here: https://www.industry.gov.au/sites/default/files/June%202021/document/australias-ai-action-plan.pdf

[16] *Singapore Model AI Governance Framework*, available here: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

8. *How organizations take into account benefits and issues related to inclusiveness in AI design, development, use and evaluation—and how AI design and development may be carried out in a way that reduces or manages the risk of potential negative impact on individuals, groups, and society.*

We discuss the concept of inclusiveness in response to Q3, as we believe this is an important foundational principle that should be included in the AI RMF. Many of our member companies are considering how to design AI with inclusivity in mind, undertaking things like user studies to understand how a person interacts with and experiences AI, making sure that the data sets used to train AI systems are representative and that models are designed with a variety of attributes in mind, and actively monitoring how an AI system performs so that adjustments can be made if necessary.

We also highlight the importance of inclusivity in facilitating public trust in and understanding of AI technology in our *Global AI Policy Recommendations* and offer several suggestions that may help to reduce potential negative impacts of AI on individuals, groups, and society.

For example, we recommend that one way in which developers can reduce potential negative impact and improve trust is to partner with universities whereby data science and other students in aligned disciplines conduct real world projects with communities in key areas of social need. Doing so can significantly improve students' skills while also providing a tangible benefit to social groups in need. Such partnerships also serve a training function for the communities involved, who learn what problems AI can and cannot solve, and how to make the technology work for them in a beneficial way.

We further recommend considering how to develop meaningfully explainable AI systems, as this is another way to reduce potential negative impacts. We appreciate that NIST has already started to undertake this work with the publication of its *Four Principles of Explainable AI*. In the context of a risk management framework, it may be worth exploring how explainability can help foster accountability, by in turn helping entities to make decisions that avoid negative outcomes. While explainability is not helpful in every instance, there may be some higher-risk use cases where it makes sense. We therefore encourage NIST to consider how explainability could play a role in risk management, keeping in mind that not all outcomes will require an explanation (and that it some cases, an explanation may not be possible).

9. *The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes");*

We strongly support all of the attributes that NIST has laid out in the RFI on the AI RMF. All of these will be important to developing a valuable risk management framework that can be leveraged by a variety of stakeholders. A consensus-driven, open process will ensure that NIST achieves buy-in from the broader community and will also ensure that the AI RMF

includes relevant guidance, approaches, and best practices from across groups. We believe that the processes that NIST has undertaken in developing the Cybersecurity and Privacy Frameworks have been undeniably successful and should be emulated here.

We also agree that the AI RMF should provide common definitions. One of the major issues we have observed in discussing AI and AI concepts with different stakeholders in the United States and around the world is that there is no common definition for many concepts, including for AI itself. Oftentimes we are talking about different things using the same terminology (e.g., transparency may equate to explainability for some, but that the AI algorithm or system was designed in an open fashion to another), so developing common definitions for key AI concepts will be immensely helpful. In doing so, using plain, understandable language will be of paramount importance. We have attempted to cull together a variety of definitions in our *Global AI Policy Recommendations,* which may be helpful to NIST as it attempts to establish definitions. It may also be helpful to leverage ISO/IEC DIS 22989 AI Concepts and Terminology, which, while not yet complete, may offer a starting point.

Beyond that, the attributes NIST outlines related to taking a risk-based, outcome-focused, voluntary, and non-prescriptive approach and ensuring consistency with other approaches to managing AI risk align with the recommendations we set forth in our *Global AI Policy Recommendations.* Those recommendations emphasize the importance of taking a risk-based approach to AI governance, where risks are identified and mitigated in the context of a specific AI use case. We also strongly believe that the AI RMF should seek to maintain consistency with other AI approaches around the globe. Political and regulatory divergence poses 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.

Finally, we strongly agree that the AI RMF should be a living document and appreciate that NIST recognizes there are aspects of AI trustworthiness that are not sufficiently ripe for inclusion in a risk management framework. In keeping with our recommendation to look beyond trustworthiness at the outset, we urge NIST to also consider that there may be AI risks or functions more broadly that are not ripe for inclusion in the AI RMF or that some AI risks remain unknown at this time. 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.

10. *Effective ways to structure the Framework to achieve the desired goals, including, but not limited to, integrating AI risk management processes with organizational processes for developing products and services for better outcomes in terms of trustworthiness and management of AI risks. Respondents are asked to identify any current models which would be effective. These could include—but are not limited to—the NIST Cybersecurity Framework or Privacy Framework, which focus on outcomes, functions, categories and subcategories and also offer options for developing profiles reflecting current and desired approaches as well as tiers to describe degree of framework implementation; and*

Because one of the challenges to developing the AI RMF is the lack of established standards, guidelines, and best practices for AI risk management, we think it might be useful for NIST, as an initial step, 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.[17] 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 likely be of more value, 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.

Similar to the NIST Cybersecurity Framework, we believe an outcomes-based approach is the most effective way to achieve the desired goals of the AI RMF. A focus on the outcomes of AI models (e.g., whether the model results are leading to discriminatory outcomes or disparate impact) will help to protect against the risks of AI while still facilitating innovation and agile development. Rather than focusing on technically prescriptive measures, which can result in slow adoption and barriers to entry, the AI RMF should incentivize proper risk management, accountability, and ethical considerations, while allowing companies to operate flexibly and efficiently and keep pace with innovation. The objective should be to mitigate risk and protect consumer privacy while promoting opportunities for innovation.

As a part of the AI RMF, we believe it would be useful for NIST to develop a methodology that can help stakeholders determine the risk-level of a specific AI use case and then take steps based on that identification to mitigate that risk. 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 and users in making risk determinations. Importantly, 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 developers and users to figure out what level of a risk a particular use

[17] *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

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.

We also believe that the AI RMF should help stakeholders determine how to navigate tensions that may arise in developing and using AI, which we have referenced as a challenge in response to Q1. For example, how does a developer balance fairness with privacy? Reducing bias or mitigating biased outcomes generally requires the collection of *more* data, which could come into conflict with protecting privacy. This is a very real tension that developers face with no real guidance as to how to resolve those sorts of conflicts. We hope that the AI RMF provides a helpful formula for weighing privacy tensions with those of delivering robust and safe AI experiences throughout the AI lifecycle. It would be helpful to address within the AI RMF how, if the output of a model improves privacy for individuals, it might be possible at the model training and evaluation stages that such models be trained on personal data, even if doing so appears to superficially conflict with privacy-by-design and data minimization principles.

As we noted earlier in our response (see *General Thoughts* p. 2), AI risk management is unique in that the contexts and applications of this technology can be very different, even when the underlying algorithm is the same. For example, a convolutional neural network can be used for collision avoidance in commercial quadcopter drones equipped with cameras. A different convolutional neural network (the identical algorithm) could be leveraged to conduct identification and surveillance on certain groups of people. The risks associated with these uses are obviously very different, even though the algorithms are identical. As such, the AI RMF should account for the following three factors:

- The deployment context
- The training data and optimization function
- The goal of the product

In some cases, it will not be possible to effectively measure the risk of AI technologies, in large part because AI is an emerging technology. In these cases, the AI RMF should not prohibit the development of AI technologies but should instead provide reasonable mitigation strategies that balance both the possible risks but also the benefits that AI technologies may offer.

11. *How the Framework could be developed to advance the recruitment, hiring, development, and retention of a knowledgeable and skilled workforce necessary to perform AI-related functions within organizations.*

We recognize the importance of furthering the talent pipeline and developing a workforce that is appropriately skilled to address AI-related functions but feel that NIST would be best served by focusing the AI RMF on risk management issues and addressing AI workforce development separately.

12. *The extent to which the Framework should include governance issues, including but not limited to make up of design and development teams, monitoring and evaluation, and grievance and redress.*

While all of the above are important issues, we encourage NIST to focus the initial iteration of the AI RMF on risk management and avoid attempting to address governance issues. As it is, the AI RMF will need to be a relatively comprehensive document in order to capture the wide variety of AI risks and potential mitigations, so adding in additional issues may serve to unnecessarily complicate the effort. Instead, it may be helpful for NIST to develop a Framework Roadmap akin to those developed for both the Privacy and Cybersecurity Frameworks. Such a Roadmap could outline design and development teams, monitoring and evaluation, and grievance and redress as priority issues for further consideration and development.

***

Once again, we appreciate the opportunity to provide feedback to NIST's RFI on the AI RMF. We believe that such a tool can 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 address challenges AI developers face, 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.

Sincerely,

John S. Miller
Senior Vice President of Policy
and General Counsel

Courtney Lang
Senior Director of Policy