# Summary Analysis of Responses to the NIST Artificial Intelligence Risk Management Framework (AI RMF) - Request for Information (RFI) National Institute of Standards and Technology (NIST)

October 15, 2021

## Introduction

The National Institute of Standards and Technology (NIST) is developing a voluntary artificial intelligence (AI) risk management framework (RMF) to improve the management of risks to individuals, organizations and society associated with AI. Specifically, the framework would help to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. NIST has been working with the AI community to identify the building blocks: the characteristics needed to cultivate trust and ensure that AI systems are accurate, explainable and interpretable, reliable, robust, safe, secure and resilient, privacy preserving – and that they mitigate harmful bias while also taking into account fairness and transparency.

NIST is developing the framework in open and transparent collaboration with public and private sector stakeholders as directed by Congress[1] and consistent with NIST's general approach of developing iterative drafts for public comment. NIST anticipates producing a completed version 1.0 of the AI RMF in early 2023.

On July 29, 2021, NIST issued a Request for Information (RFI)[2] seeking input from stakeholders on the development of the AI RMF. As of the date of this publication, NIST has received 106 responses from a range of stakeholders, including individuals and organizations representing industry, government, and the broader public interest.[3] The responses were supportive of NIST's effort to develop the AI RMF.

Figure 1 illustrates a distribution of the sectors responding to the RFI. Responses provide a helpful starting point for discussions at the October 19-21, 2021, "Kicking off NIST AI Risk Management Framework" Workshop. The AI RMF will be developed through a consensus-driven, open, and collaborative process that will include additional workshops and opportunities for stakeholders to provide input that will be used to help inform, refine, and guide the development of the framework.

---

[1] H. Rept. 116-455—COMMERCE, JUSTICE, SCIENCE, AND RELATED AGENCIES APPROPRIATIONS BILL, 2021, CRPT-116hrpt455.pdf (congress.gov), and Section 5301 of the National Artificial Intelligence Initiative Act of 2020 (Pub. L. 116-283), https://www.congress.gov/116/bills/hr6395/BILLS-116hr6395enr.pdf

[2] Federal Register Notice 86 FR 40810, Artificial Intelligence Risk Management Framework, https://www.federalregister.gov/documents/2021/07/29/2021-16176/artificial-intelligence-risk-management-framework; Notice of Extension: Federal Register Notice 86 FR 47296, Artificial Intelligence Risk Management Framework, https://www.federalregister.gov/documents/2021/08/24/2021-18108/artificial-intelligence-risk-management-framework

[3] The responses are posted at: https://www.nist.gov/itl/ai-risk-management-framework/comments-received-rfi-artificial-intelligence-risk-management

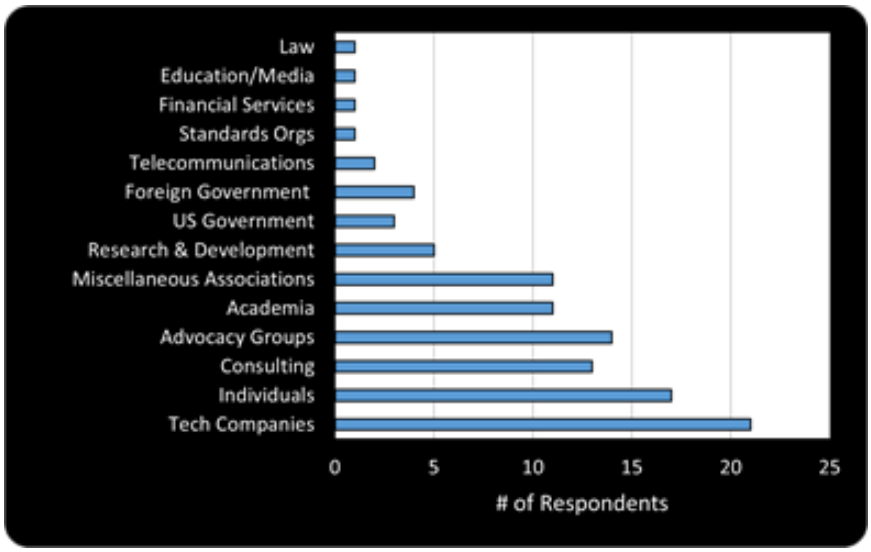


*Figure 1: RFI Respondents by Sector*

The following section explains the methodology NIST used to analyze the RFI responses and describes major themes that emerged.

## Analysis Methodology

Each RFI response was reviewed and analyzed by NIST according to the following process:

- Determine basic information about respondents to assess coverage of responses across sector and organization type;
- Associate sections of RFI responses to RFI topics; and
- Identify themes from key points and reoccurring language across responses.

For this summary analysis, NIST has focused on the RFI responses that provided information relevant to the development of the AI RMF. While some responses included information on other topics, considerations not directly related to framework development are not included in this report.

The resulting themes, described in the next section, will inform discussions at upcoming AI RMF development workshops and in NIST's development of approaches for the AI RMF.

## Key Themes from the RFI Analysis

NIST appreciates the many responses received which provide helpful insights and observations regarding development and implementation of the AI RMF. Key themes that NIST observed are described below, each with example excerpts drawn from the responses. The cite for each quotation is available through endnotes. Quotations selected for inclusion in this report are for illustrative purposes only; there may be other responses that addressed the theme as well.

## — Theme: Defining Common AI Terminology for Communicating Risk

Numerous respondents highlighted the need for plain-language, commonly adopted definitions for many of the terms. Respondents pointed to some sources of definitions that already exist (e.g., OECD) and suggested it might be beneficial for the AI community to either codify or adjust terminology. While there is broad consensus on the need to manage risk in an AI context, it was suggested that there is an opportunity for broader definition of a taxonomy of risk factors (and their relationship to other enterprise risk categories). A key consideration that surfaced in a number of comments is that risk analysis itself could be biased, so additional guidance on how to conduct risk assessment in a fair and accurate manner might be a helpful outcome.

- "The Framework should include definitions and templates that facilitate information sharing about AI risks and incidents. … Standardized ways to share information about incidents would be very valuable for identifying, assessing, prioritizing, mitigating, and communicating AI risk."[1]
- "NIST should strive to harmonize definitions of key terms with those already published so that the global AI community is speaking the same language. A common lexicon will give organizations and society more confidence and promote greater alignment of standards, frameworks, models, etc."[2]
- "We feel that it is vital for the RMF to place a particular focus on defining standard terminology around such topics as performance measurement, traceability, interpretability, explainability, transparency, repairability, and so-on, in a specific and technically relevant manner."[3]

Some respondents shared concern about low-probability, very-high impact events (sometimes referenced as a Black-Swan Event). These events are often missing or underrepresented in Machine Learning training data sets. Identification of risk factors that could lead to such events could be included as part of the overarching risk communication approach, they suggested.

- "AI systems could pose risks of catastrophe from malicious or unintentional misuse, accidents, or other failures. Posner generally uses the term catastrophe to mean "an event that is believed to have a very low probability of materializing but that if it does materialize will produce a harm so great and sudden as to seem discontinuous with the flow of events that preceded it" (Posner 2004, p. 6). Bostrom and Ćirković (2008, pp. 2-3) define global catastrophic risks as risks of serious events (e.g., with millions of fatalities or trillions of dollars of economic loss) with global scale."[4]

## — Theme: Recognize the Impact of AI on Society

Several respondents noted impacts to society and described a need for risk mitigation approaches to address such concerns. A majority of respondents who commented on AI impacts to society indicated the need for risk assessments or risk impact assessments. A few

respondents pointed out specific high risk sectors and organizational processes that may cause societal harms. Some respondents are looking for actionable guidance for how to reduce the risk of such impacts. Suggestions for reducing or mitigating negative impacts to society included use of cross-functional teams, AI-specific training for different audiences, use of an ethics board, audit and accountability practices, examining privacy protections, improved explainability and interpretability for AI systems, and examining data drift in AI model training.

- "Safe, ethical, and effective AI systems can provide tremendous benefits to society. However, we understand that AI is increasingly being developed and deployed within critical processes (e.g., healthcare, employment, judicial, policing, etc.) where there is a concern that such systems could pose a risk to safety, privacy, and human rights."[5]
- "The public's trust in AI will be eroded by widespread and unmitigated discrimination and discriminatory impact. Communities are looking for innovation that reverses racial inequality. To cultivate the public's trust in the design, development, use, and evaluation of AI technologies, AI developers and AI users must detect and remove bias from AI. Proactive algorithmic accountability will ensure that communities are receiving fair and accurate results."[6]
- "In order for AI design and development to be carried out in a way that reduces the negative impact on individuals, it is critical that AI algins with human values and norms. ... Human society will need to enact guidelines, policies, and regulations that can address issues raised by the use of AI systems, such as ethical standards that regulate conduct. These guidelines must take into account the impact of the actions in the context of the particular use of a given AI system, including potential risks, benefits, harms, and costs, and will identify the responsibilities of decision makers and the rights of humans."[7]
- "Companies that practice Responsible AI—and let their clients and users know they do so—have the potential to increase market share and long-term profitability. Responsible AI can be used to build high-performing systems with more reliable and explainable outcomes. When based on the authentic and ethical strengths of an organization, these outcomes help build greater trust, improve customer loyalty, and ultimately boost revenues."[8]

## — Theme: Methods for Structuring the AI RMF

Recognizing the evolutionary nature of AI, responses indicate support for a continuous process that enables collaboration and communication among workforce, risk management, and enterprise strategy stakeholders. Respondents indicate support for a framework structure that documents organizational goals and applies an outcome-based approach to achieve them.

- "Positioning the framework as a continuous learning process can help to introduce the notion that everyone has a role to play in learning about the evolution of AI systems, the risks that emerge, and strategies for addressing them. By focusing the framework on

learning toward the desired systems outcomes (i.e., systems that are trustworthy, secure, resilient, etc.) it broadens the aperture to include multiple approaches for how to reach end states, rather than focus on a single approach adopted by individuals with fixed roles and skills."[9]

- "We strongly support development of an AI RMF that provides a catalog of outcomes and approaches applicable for a variety of use cases, rather than a set of one-size fits-all requirements, considering that:
  - the catalog supports the prioritized, scalable and cost-effective objectives,
  - the rigor and sophistication of AI risk management should be commensurate with the impacts of AI system outcomes to individuals, groups, society and organizations, and
  - the relevance of the principles and characteristics for AI trustworthiness significantly varies depending on its intended use."[10]
- "We believe the proposed RMF should not stand alone but should be included within a general framework for AI engineering that includes not only AI design principles and best practices, but also guidelines for robust, fair, and ethical use and performance. In this way, risk consideration can more closely follow the AI lifecycle throughout rather than being an afterthought."[11]

### Sub-Theme: Alignment with Non-AI Risk Management Resources

Some respondents noted that many organizations have made significant investments applying existing non-AI risk management resources and that existing frameworks, standards, guidelines, methodologies, and tools from NIST, other SDOs (standards developing organizations) and industry partners may be helpful in developing the AI RMF. Respondents indicated that the AI RMF should, to the extent possible, be aligned with these approaches, while recognizing the significant and unique differences between AI risks and other enterprise risk factors.

- "We encourage NIST to leverage principles already incorporated into other frameworks such as the NIST Cybersecurity Framework (CSF) and the NIST Privacy Risk Management Framework, as well as the five principles embodied in the Committee of Sponsoring Organizations (COSO) Framework: governance; strategy; performance; review & revision; and information, communication & reporting."[12]
- "We encourage NIST to drive for alignment between definitions provided in the AI RMF with definitions proposed in the European Commission's proposed AI Act as well as definitions under development at the International Standards Organization."[13]
- "An important part of this will be ensuring the AI RMF is informed by the significant amount of work that has been done in recent years to create common standards and frameworks for identifying and mitigating AI risk. This will include reflecting and driving alignment around existing international standards, including the many existing technology

standards that are applicable to AI even if they were originally developed for other technology segments, including data formats, transfer protocols, cybersecurity practices, privacy practices and cloud services practices."[14]

Few respondents mentioned that a noteworthy aspect of AI risk is that, due to continuous learning of an algorithm, once a model has been put into production, accuracy may be reduced as more real-world data is introduced. In some cases, model retraining happens automatically, so changes in risk factors may not be immediately observable. They added that these examples illustrate that the AI RMF will need to work with existing risk methods, but that there are unique and specialized processes that will need to be adopted for AI.

### Sub-Theme: Correlate AI Risk Management with an AI System Life Cycle

The value of a life cycle approach resounded throughout the comments received. Several respondents described the need for a clearly defined and agreed-upon approach, aligned to other organizational processes. Inclusion of an AI-specific life cycle will highlight areas where AI risk is unique from other risk types. They suggested that doing so also would help with risk assessment, identification of requirements and gaps, and may support collaboration with internal and external stakeholders.

- "A process that is applied at all stages of the AI engineering lifecycle ensuring that any intelligent system is producing outcomes that are valid, verified, data-driven, trustworthy, and explainable to a layman, ethical in the context of its deployment, unbiased in its learning, and fair to its users."[15]
- "Recognizing that the specifics of a governance program will necessarily vary according to the size and capacity of organizations, NIST should consider their basic elements. These include the involvement of senior management, such as appropriate C-Suite executives, to oversee the company's AI product development lifecycle, and a trustworthy AI compliance team responsible for carrying out impact assessments, documentation, training, and serving as a cross-company resource."[16]
- "AI systems are dynamic, with many continuing to adapt throughout their lifecycle as they learn from the data they process, as are the societal contexts into which they are deployed. Moreover, AI is developed, operated, and maintained as a service rather than as a fixed product. As such, it is important to adopt a lifecycle approach to monitoring and responding to the risks of a particular deployment given the way its performance may change over time. Assessment of risk over the lifespan of an AI system must address changing behaviors, workloads, and associated outcomes due to both (1) updates to data, models, parameters, and overall functionality of deployed systems that may come via maintenance and updating, and (2) changes in the nature or distributions of tasks or workloads analyzed or handled by the system over time."[17]

## — Theme: Considerations for AI Governance

Throughout the comments, respondents highlighted the need for NIST to provide guidance regarding how to achieve an overarching governance system for protection of AI systems and for identifying, assessing, and monitoring for potential harms. The components of such a governance system would identify stakeholder goals (e.g., risk appetite and risk tolerance), compliance needs, other operational requirements, and oversight capabilities.

» "Entities should implement governance structures for AI systems that incorporate organizational values, consider risks, assign clear roles and responsibilities, and involve multidisciplinary stakeholders. Entities should include diverse perspectives from technical and non-technical communities throughout the AI life cycle to anticipate and mitigate unintended consequences including potential bias and discrimination."[18]

» "AI governance frameworks could help organizations learn, govern, monitor, and mature AI adoption. The four core components of AI governance identified in the white paper are definitions, inventory, policy/standards, and a governance framework, including controls."[19]

» "We strongly recommend that the Framework include a comprehensive set of governance mechanisms to help organizations mitigate identified risks."[20]

Many of these comments focused on organizational governance, though the need for data governance (i.e., the practices and processes to ensure the management of data within an organization) was observed, as well. Because of the connection between the data used to train and create algorithms, data governance (including data stewardship and data quality) is likely to be highly related to AI risk discussions. One respondent shared that, "trust in the design, development, and use, of AI requires that the data upon which AI is built and trained be subject to a rigorous data governance program and that the key tenets and procedures of that data governance program be transparent to both developers of AI and their end users."[21]

Several subsidiary themes related to governance were observed, including those below.

### Sub-Theme: Incorporating Ethical Principles into Governance

Many respondents commented on the need for AI ethics as a system of principles and techniques for the development and proper use of AI technology throughout the AI life cycle. Ethics, as part of a governance structure, may help to both establish and maintain AI risk management processes that align with organizational values and goals. The framework will need to address how best to include references to ethics – for example, potentially as "AI ethical principles" to inform and guide management of AI risk, for example, as opposed to "ethical AI products," "ethical AI services," or "ethical AI technologies."

» "Ethical AI is about values driving application, and any organization considering how to manage, design, evaluate and use AI must start by developing guiding principles anchored

to mission driven core values. By aligning AI principles to values, organizations can create a positive impact and reduce unintended harm."[22]

- "AI ethics should be centered around human ethics. A collaborative effort is necessary to determine what ethical values society wants AI to reflect. These should be centered on core principles that are defined and mapped to the ethical framework. While the risk framework is developed separately, we hope that NIST will connect this product to ethical principles developed with the input of AI stakeholders."[23]
- "As with technology, ethics does not emerge in a vacuum and does not on its own determine sociotechnical outcomes. What is needed, then, is a new approach that looks to the sociotechnical complexities of tech ethics: how it shapes the development and governance of technologies (and collective understandings of technologies) as well as how ethical discourses and practices are themselves shaped by a variety of social forces."[24]

### Sub-Theme: Clear Definition of Relevant AI Workforce Roles

Many respondents expressed that a key element of governance is the composition of the AI workforce and determination of appropriate roles and their associated responsibilities. In addition to the need to define the workforce, many respondents recommended incorporating a diverse and inclusive workforce to reduce risk and improve efficiency. They supported an AI workforce with the skills and training to engage the framework successfully.

- "By showing what the key risks and controls are for AI/ML, what is needed to mitigate them, and what skillsets and job functions are best suited to each role in the process, NIST can greatly help organizations build proper AI/ML development and risk management functions, as well as help to alleviate the ambiguity that exists for many organizations."[25]
- "A diverse workforce with a broad perspective and understanding of risks associated with AI applications is necessary to identify, prioritize, and respond to risks. The challenge of creating a diverse workforce for AI requires a holistic approach, starting from early education and throughout a career. It must focus not just on recruiting talent but also on developing and retaining existing talent, which requires looking at an organization's culture and whether it is inclusive."[26]
- "Studies have shown that one of the most important determinants of a team's ability to confront issues of harmful bias in AI is that team's diversity. Less diverse teams have a harder time reducing unintended bias in their machine learning models than teams that are made up of members that come from a wide range of genders, ethnicities, and backgrounds. Furthermore, teams that do not represent the perspective of communities impacted by AI systems have a harder time predicting and mitigating potential harms that the system causes to those communities."[27]

### — Theme: Metrics and Monitoring of AI Systems and Data

Many respondents noted the need for consistent metrics and other means of measuring and communicating performance regarding the achievement of organizational goals. Metrics could support not only measurement of performance regarding risk management strategy, but also enable monitoring to detect data governance risks (e.g., conscious and unconscious bias within processes).

- "Entities should document requirements—including performance metrics—for the AI system throughout the life cycle." "Entities should document methods to assess performance—which can include input-output tests, stress tests, and evaluations of model drift—to ensure AI systems meet their intended goals." "Entities should provide access to performance test results, change logs, and other documentation describing updates and key design choices, and provide a copy of the model or algorithm code to third-party assessors of AI systems."[28]
- "Current understanding of...performance measures, particularly outside of isolated testing of individual components, is poorly understood. It is strongly recommended that the RMF provide guidelines for the initial and continued performance measurement of AI systems as part of whole, and interconnected, systems."[29]
- "Currently, the greatest challenge in AI-risk management policy is the inability to anticipate model failure. Until the AI community defines a comprehensive list of model susceptibilities, from gender and racial bias to adversarial examples and manipulation to bad actors, this weakness is impossible to measure and improve upon. Therefore, it is imperative to identify a comprehensive framework or schema of what could go wrong in a model's performance, and then test rigorously against these factors."[30]

## Advancing the Dialogue

The comments received reinforced that there is a significant amount of foundational work on which to build AI RMF discussions – and a variety of views about what constitutes risk and trustworthiness. The responses reflected the challenge that risk management discussions often entail establishing the right balance among the benefits of information and technology while managing the risk and resource implications. In the case of the AI RMF, NIST also received very specific concerns and suggestions about managing AI risks related to civil rights, civil liberties, and equity. Maximizing effectiveness through reuse of existing methodologies will help to accelerate further adoption of AI risk management, but these considerations must be balanced with organizations' understanding that there are unique and challenging aspects to AI risk, including potential direct individual and societal harms. As AI solutions continue to proliferate, the need for effective balance also increases. NIST encourages a broad and diverse dialogue on these issues and participation throughout the development of this important framework.

1 Center for Security and Emerging Technologies, at 2, https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0100.pdf

2 U.S. Chamber of Commerce's Technology Engagement Center, at 4, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0084.pdf

3 Raymond Sheh and Karen Geappen, at 1, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0093.pdf

4 U.S. Chamber of Commerce's Technology Engagement Center, at 3, https://www.nist.gov/document/ai-rmf-rfi-comments-us-chamber-commerces-technology-engagement-center-ctec

5 UC Berkeley, at 7, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0092.pdf

6 Color Of Change, at 7, https://www.nist.gov/document/ai-rmf-rfi-comments-color-change

7 Computing Community Consortium, at 3, https://www.nist.gov/document/ai-rmf-rfi-commnets-computing-community-consortium-ccc

8 Boston Consulting Group, at 2, https://www.nist.gov/document/ai-rmf-rfi-comments-boston-consulting-group

9 Rachel Dzombak et al, Carnegie Mellon University at 7. https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0085.pdf

10 Takashi Suzuki, Blackberry at 1. https://www.nist.gov/system/files/documents/2021/09/14/ai-rmf-rfi-0060.pdf

11 Kendra Kratkiewicz et al, MIT Lincoln Laboratory at 2. https://www.nist.gov/system/files/documents/2021/09/22/ai-rmf-rfi-0089.pdf

12 Matt Baker, Deloitte Government and Public Services, Deloitte & Touche LLP, at 1, https://www.nist.gov/system/files/documents/2021/09/15/ai-rmf-rfi-0073.pdf

13 Unity, at 4, https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0095.pdf

14 Microsoft Corporation, Microsoft, at 6, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0088.pdf

15 Laura Freeman and Feras A. Batarseh, Virginia Tech, at 4, https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0097.pdf

16 Workday, at 5, https://www.nist.gov/document/ai-rmf-rfi-comments-workday-attachment-1

17 Microsoft, at 2, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0088.pdf

18 Dennis H Mayo, PhD , Government Accountability Office. https://www.nist.gov/document/ai-rmf-rfi-comments-dennis-h-mayo-phd-government-accountability-office

19 Yogesh Mudgal, Artificial Intelligence Risk & Governance, Wharton, at 2, https://www.nist.gov/system/files/documents/2021/08/19/ai-rmf-rfi-0022-attachment.pdf

20 UC Berkley, at 6, https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0092-attachment.pdf

21 Mary Kathryn Rondon, at 2, https://www.nist.gov/system/files/documents/2021/09/14/ai-rmf-rfi-0061.pdf

22 Booz Allen Hamilton at 2. https://www.nist.gov/system/files/documents/2021/09/13/ai-rmf-rfi-0050.pdf

23 Bipartisan Policy Center at 4. https://www.nist.gov/system/files/documents/2021/08/20/ai-rmf-rif-0026.pdf

24 Susan von Struensee at 25. https://www.nist.gov/document/ai-rmf-rfi-comments-ben-green-michigan-society-fellows

25 Monitaur, Inc. at 6. https://www.nist.gov/system/files/documents/2021/08/27/ai-rmf-rfi-0054.pdf

26 Bipartisan Policy Center at 2. https://www.nist.gov/system/files/documents/2021/08/20/ai-rmf-rif-0026.pdf

27 Singh, Eisenberg, Shattuck, and Davidson at https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0104.pdf

28 Dennis H Mayo, PhD , Government Accountability Office. https://www.nist.gov/document/ai-rmf-rfi-comments-dennis-h-mayo-phd-government-accountability-office

29 Raymond Sheh and Karen Geappen at 4. https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0093.pdf

30 Corner Alliance at 1. https://www.nist.gov/system/files/documents/2021/08/19/ai-rmf-rfi-0017.html