



NOVEMBER 2023

Operationalizing the NIST AI RMF Framework

Introduction

Artificial Intelligence (AI) is already impacting many aspects of people's lives and society. At Carnegie Mellon University (CMU), we believe that AI systems must be designed, developed, and deployed responsibly to ensure accountability and transparency, and promote a more just and equitable society. The Block Center's Responsible AI initiative brings together the university's cutting edge educators and researchers and their expertise in partnership with public and private sector experts to advance effective policy making and practical knowledge, generate thought leadership, and contribute to the timely discourse around the responsible use of AI. One such public sector collaborator is the National Institute of Standards and Technology (NIST) within the U.S. Department of Commerce. NIST is responsible for the development of measures and standards for emerging technologies in the U.S.

AI continues to emerge as one of the most important developing sectors of the American technology industry, with far reaching implications for both economic and societal development. In response, in January 2023, NIST published its AI Risk Management Framework (AI RMF), a document developed in partnership with public and private partners (including faculty members from CMU and the Block Center), to help organizations better manage and mitigate the risks associated with AI (NIST 2023). NIST intends to enhance the ability of developers to include trustworthiness and responsibility into the design, deployment, and evaluation of AI systems. It provides guidelines for identifying the potential risks in the use of AI, including potential harms to people, organizations, and ecosystems; while simultaneously providing guidance on incorporating trustworthiness into system design to help mitigate these potential harms ([NIST 2023](#)). The AI RMF consists of four core functions; Govern, Map, Measure, and Manage, each with a series of categories and subcategories for interested parties to utilize as they design and deploy their AI system or product (NIST 2023).

In July 2023, the Block Center and NIST convened a workshop at CMU on how to operationalize the use of the AI RMF across different sectors. Experts from across the country met to discuss and present on how to operationalize the framework through potential use-cases. The first day featured a round-table discussion on potential AI RMF use-cases across different sectors, including financial services, labor, healthcare, local government, and non-profit/community organizations. The second day of the event consisted of a series of discussions on various elements of AI governance and deployment. This white paper is intended to synthesize and provide an overview of the key four themes of these discussions that are critical in helping organizations operationalize the AI RMF: (1) evaluation of AI systems through a socio-technical lens; (2) creating public sector procurement guidelines; (3) enhancing and diversifying potential stakeholder engagement; and (3) providing education and training in responsible AI.




The NIST AI RMF and Potential Use Cases

The AI RMF is composed of four core functions: an overarching Govern function and three branch functions — Map, Measure, and Manage. Each of these core functions consists of a breakdown of categories and subcategories meant to assist stakeholders in introducing best practices for risk management in the development and deployment of AI products.

- The **Govern function** introduces an overarching series of standards and practices meant to enhance stakeholder risk management measures and permeate each step of AI development to enable the Map, Measure, and Manage functions.
- The **Map function** allows stakeholders to *identify and frame* risk within an AI product, establishing the context of the product's functions and categorizing risk and benefits associated with these functions. These identification activities provide necessary data for the observation and measurement of risk found in the Measure function.
- The **Measure function** uses a variety of techniques to *analyze AI risk and their potential impact*. This process includes using both quantitative and qualitative data to assess AI risks and subsequently provides information for mitigation and monitoring activities of the Manage function.
- Finally, the **Manage function** determines the *appropriate risk management response* given the data provided by the mapping and measurement activities, using the information to prioritize risk mitigation and management activities.

We, at the Responsible AI Initiative, strongly believe that processes and tools for operationalization need to be grounded in real-world use cases and done in partnership with on-the-ground organizations from the public and private sector. The workshop brought a selection of those organizations to CMU to 1) describe their use cases and 2) brainstorm with researchers and educators around how to support them in operationalizing the AI RMF, maximizing the positive impact of AI while guarding against potential negative impacts.

Human Services

Allegheny County's Department of Human Services (DHS) utilizes data and AI tools to allocate its limited resources through its Homeless Support System. The county only has access to 800 beds for its homeless population of close to one thousand individuals. In 2020, DHS launched a new approach to updating its model, which previously relied on self-reporting of information and used an outdated data system that DHS employees struggled to navigate. The new model includes fairness reviews from third-party evaluators and includes more advanced data capabilities to improve both mental and physical health outcomes for the sheltered homeless population. The success of this new housing support model relies on the regular monitoring of the tool, and consistently loops human operators into the process to enhance its reliability and transparency. Both the third party fairness review and the human/tool interface demonstrate the potential use of the AI RMF Governance function, ensuring that best-practices for the trustworthiness and risk-mitigation are followed throughout the development and use of an AI system/product. 



The NIST AI RMF and Potential Use Cases *(continued)*

Labor and Workshifting

Within the hotel industry, organizing the housekeeping needs and preferences of guests can be difficult to manage given the variety of guest preferences (e.g., early departure, afternoon cleaning as opposed to morning cleaning) and the potentially large number of guests. As a result, hotels are beginning to integrate AI systems to help manage their housekeeping operations, but there are unintended consequences from using these systems. AI systems do not always take into account the needs and safety of the housekeepers themselves when creating the housekeeping assignments. They also do not account for the experience or seniority of housekeepers, and housekeepers do not receive the same respect for their experience or expertise that they are accustomed to receiving when an individual is responsible for creating the assignment list. Therefore, AI systems can shift labor burdens onto different people in unforeseen ways. These workshifts demonstrate a clear potential application of the AI RMF, by mapping out the potential risk of these types of AI systems, measuring the severity of that risk, and implementing a management plan to help address the concerns of housekeepers who feel disrespected by the automated AI planning system.

Financial Services

The banking and financial services industry is considering how to use more advanced data analytics and AI systems in order to create credit score proxies for the millions of Americans who do not have adequate access to banking services or credit. There are a number of constraints that the industry faces for creating a reliable system to increase access to these financial services, including the lack of or poor quality of existing data on the unbanked population in the United States, the limited number of reliable proxies, and the outdated algorithms currently in use. The banking industry is interested in using more advanced data analysis techniques to help increase access to financial services, but they want to improve the trustworthiness of any AI system they use and plan for potential risks or biases that could be associated with data regarding the unbanked population. Operationalizing the NIST AI RMF will enable banks and other financial service institutions to ensure that operationalizing AI in the finance domain will be impactful, while providing a level of transparency and trustworthiness in their practices to demonstrate the responsible use of this technology.

City Planning

City planning presents another important use case for the NIST AI RMF. City planners across the country endeavor to enhance their efficiency and effectiveness by integrating AI into their operations. For example, cities can and have begun adopting AI systems that **monitor the condition of a city's paved road infrastructure**, which then helps civil servants triage and prioritize road maintenance projects. However, city employees noticed a trend where the AI tool they were using tended to recommend and





The NIST AI RMF and Potential Use Cases *(continued)*

prioritize streets in the city's wealthiest neighborhoods and the developers were unwilling to share their proprietary code. Utilizing the AI RMF in the development of these tools can identify such equity and transparency risks, and minimize their negative impacts.

Using an outside vendor to review and **trim police body camera footage** to save video storage space presents another use case for the AI RMF. If this vendor is using an AI system, there need to be checks in place to ensure that the AI is not removing events from the footage that should have been included. Considering the potential risks of using such a system and the tremendous consequences it could have are vital to the responsible deployment of AI systems in our society.

Key Themes

Four key themes emerged over the course of the event in July: (1) evaluating AI systems through a sociotechnical lens; (2) creating public sector procurement guidelines; (3) potentially enhancing stakeholder engagement; and (4) providing education and training in responsible AI for various stakeholders. Each of these themes highlight the importance of continued collaboration between thought leaders, industry experts and practitioners, and AI end-users as we continue to evaluate and promote tools like the NIST AI RMF into every-day best practice.

Evaluating AI systems through a sociotechnical lens:

AI systems are used in socially complex contexts. The human interface is a vital component of any AI system, and the analytical outputs of these systems can have far-reaching implications for communities and society at large. Given this reality, it is vital to evaluate AI systems within the social context in which they exist, and ensure that proper risk management strategies are used to maximize the good and mitigate the harm that AI systems can do for our society. AI systems need to be developed and evaluated with a human-centered approach that accounts for the social context of their use to assess fairness and efficiency.

This assessment can take several forms, including engaging with impacted communities and stakeholders during the design and development process to create a more nuanced understanding of the impacts that systems can have on these affected stakeholders. Developers need to consider important questions as they develop AI systems including: how to explain the expected impacts of a proposed system to stakeholders; how to effectively elicit feedback from impacted stakeholders; how to incorporate unstructured feedback in a way that system developers can effectively use; and how to incorporate feedback into their AI system and communicate the limitations of the feedback to the impacted stakeholders. When considered in combination, these questions will amplify the voices of impacted communities and stakeholders, providing developers with greater context and perspective to create trustworthy and effective AI systems for social good.





Key Themes *(continued)*

Creating public sector procurement guidelines:

Government procurement provides American society with a built-in check-point for the development of risk management and trustworthiness in AI systems as they advance and become more commonly used in the public sector. Government contracting at the local, state, and federal level requires robust guidelines and requirements before any product or service can be purchased by agencies. Therefore, there is an opportunity to build risk management frameworks and responsible AI development practices into the requirements for government purchase, which is a key source of capital for the development and refinement of emerging technologies. The United States Government has tremendous spending power, but strict rules for access to that capital. It is, therefore, a logical place for specific guidelines and requirements for AI products and services.

A key challenge in procurement is the purchase of off-the-shelf AI products, which often function as 'black boxes' with proprietary models, limiting the ability to externally validate and assess these systems for bias or other issues. There is an opportunity to use procurement guidelines and regulations to ensure that safe and responsible risk management practices are being used by AI developers before they can have access to government purchasing power. A potential mitigation strategy is the use of clear language and terminology in procurement office contracts. One idea is to specifically call for the demonstrable use of risk management and trustworthiness measures in the "off-the-shelf" AI products purchased and used by every level of government agency. There are examples of this concept already in use at the federal level, such as the [Defense Innovation Unit](#)'s vetting process. This vetting process could serve as a potential model for responsible AI federal procurement guidelines.

Enhancing potential stakeholder engagement:

Engagement with potential AI stakeholders will be crucial as AI technology becomes more prevalent and prolific throughout the workforce and across the public and private sectors. Individuals who do not consider themselves "tech savvy" enough to learn about the responsible use of AI need innovative engagement strategies to provide feedback on the evaluation of AI and machine learning systems. Given the highly social context of AI systems discussed above, it is important that educational toolkits are capable of explaining to the public the social impacts of machine learning models and the importance of diverse social values in AI development.



Key Themes *(continued)*

Throughout the event, participants introduced methods to help demonstrate potential innovative strategies to create public engagement and discourse. These included:

- a set of role-playing exercises intended to foster interdisciplinary thinking among students evaluating AI/ML systems;
- an AI Lifecycle Storyboarding method that allowed communities to share their input on the AI development lifecycle through the creation and discussion of comic stories;
- and a Value Card Exercise designed to facilitate conversations around concepts of fairness, accountability, transparency, and ethics related to sociotechnical systems.

Finally, participants demonstrated a web-based tool that is being developed to support community-centered AI evaluations, enabling stakeholders to contribute to the AI development lifecycle. Such tools will continue to grow in importance as AI technology becomes more prevalent across our society.

Providing education and training in responsible AI for various stakeholders:

Practical, context-aware education and training in AI systems will be vital to continued development of responsible and trustworthy AI systems. As AI technology continues to develop and become more accessible to society at large, there needs to be a better understanding of what exactly AI and machine learning systems are and their capabilities. This includes the terminology that industry experts use, as there is an unnecessary amount of confusing and contradictory language that dissuades the public from learning more and engaging in these important tools. Fluency in AI terminology is one of the most important emerging skills needed at every level of the workforce, from entry level positions to the executives who want to use advanced data analysis to improve efficiency.

There is a growing need for more practical experience and networking on data and AI skills, especially in the public and nonprofit sectors where resources are often limited. The workforce will need greater access to practical education and experiences in how to use data and AI systems, especially in identifying and utilizing appropriate data sources for AI systems. Building out these skills across the workforce will help enhance the use of responsible best-practices in the development of AI systems and products. Deepening the understanding of AI and the potential risks associated with the use of data will ease the burden of introducing guidelines and frameworks such as the NIST AI RMF across the AI industry.



Conclusion

As AI's influence expands and becomes ubiquitous, ensuring its responsible integration is not just a best practice—it's an imperative for our shared future. Thus, continued conversations about the importance of guidelines such as the NIST AI RMF will be of vital importance, enhancing the ability of developers to embed trustworthiness and responsibility into the design, deployment, and evaluation of AI systems. We must implement a set of measures and practices to operationalize these guidelines.

First and foremost, we must learn to evaluate AI through a sociotechnical lens, understanding its profound interplay with human society. Once our vision is clear, we must create public sector procurement guidelines, expand stakeholder engagement across the AI lifecycle, and develop educational materials that elucidate the context and considerations related to sociotechnical systems. Equipped with insights surrounding these measures and practices, the Block Center's Responsible AI Initiative is poised to confront the challenges and seize the opportunities AI presents. Moving forward, the Responsible AI Initiative will continue to convene cross-sectoral leaders to highlight and operationalize solutions to society's greatest challenges surrounding AI systems and products.

This policy brief was drafted and published by affiliated faculty and staff of the [Responsible AI Initiative](#) of the [Block Center for Technology and Society](#) at Carnegie Mellon University in coordination with the National Institute of Standards and Technology (NIST).

If you would like to schedule a briefing with Carnegie Mellon faculty experts to discuss AI accountability, toolkits to address bias in AI systems, or a deeper dive on policy ideas conveyed in this memo, please contact responsibleai@andrew.cmu.edu.

Co-authors listed in alphabetical order: [Jodi Forlizzi](#), Cole Gessner, [Rayid Ghani](#), [Hoda Heidari](#), Harrison Leon, [Steve Wray](#).