MORRISON FOERSTER

Risky Business: NIST Releases New AI Risk Management Framework

16 Mar 2023

Privacy + Data Security

Client Alert

The National Institute of Standards and Technology (**NIST**) published its **AI Risk Management Framework** (**RMF**) on January 26, 2023, to assist organizations with managing current and future risks associated with AI. The RMF was published pursuant to a congressional mandate set forth in the National Artificial Intelligence Initiative Act of 2020 and is the culmination of several NIST drafts and input by hundreds of stakeholders.

The RMF sets forth voluntary guidance for organizations to use in designing, developing, deploying, and using AI products, services, and systems (each, an "**AI System**" and, collectively, "**AI Systems**") and encourages organizations to think about, discuss, and monitor their AI Systems and their potential positive impacts and risks. The RMF provides a measurable process to enable organizations to consider and address potential issues. As the AI landscape continues to develop, NIST plans to actively update, improve, and expand on the RMF, based on community feedback, as well as the general development of AI technologies, practices, and standards globally. While the RMF provides voluntary guidelines, NIST publications often become industry standard practices and may be used as the basis for evaluating processes during due diligence reviews and regulator inquiries. For example, the Federal Trade Commission (**FTC**) stated in a **2016 blog post** that the core functions detailed in NIST's voluntary Cybersecurity Framework corresponded with "alleged lapses" challenged by the FTC in its law enforcement actions.

The RMF outlines **four key functions** intended to serve as actionable steps for organizations in their evaluation and use of AI Systems:

1. **Govern** is meant to aid with cultivating and implementing a culture of risk management within organizations designing, developing, deploying, evaluating, or acquiring AI Systems. It outlines the processes, documents, and organizational schemes involved in anticipating and managing the risks that an AI System can pose. Additionally, this function provides for procedures under which AI risk management functions can align with organizational principles, policies, and strategic priorities.

2. **Map** enhances an organization's ability to proactively identify contextualized risks posed by AI Systems. This enables risk prevention and informs preliminary decisions about the relevance or the necessity for AI Systems. The Map function is meant to serve as a basis for the Measure and Manage functions.

3. **Measure** utilizes quantitative and qualitative tools and techniques to analyze, assess, benchmark, and monitor risks posed by AI Systems. Outcomes from the Measure function are utilized in the Manage function to help with risk monitoring and response efforts.

4. **Manage** involves allocating risk resources to mapped and measured risks on a regular basis. The goal of the Manage function is to plan how to effectively respond to, recover from, and communicate about incidents or events related to AI Systems. The Manage function provides strategies to maximize AI benefits and ensures that response, recovery, and communication plans for identified and measured AI risks are regularly documented and monitored.

The RMF also identifies the following characteristics of trustworthy AI Systems:

- **Valid and Reliable:** AI Systems are often assessed by ongoing testing or monitoring to confirm that the systems are performing as intended. Validity and reliability measurements contribute to the trustworthiness of AI Systems.
- **Safe:** Safety of AI Systems is rooted in responsible design, development, and deployment practices, as well as responsible decision-making by deployers and end users.
- **Secure and Resilient:** Resilient AI Systems can return to normal function after an unexpected or adverse event. Secure AI Systems are those that can maintain confidentiality, integrity, and availability through mechanisms that prevent unauthorized

Contacts

Marian A. Waldmann Agarwal
mwaldmann@mofo.com
(212) 468-7900
(212) 336-4230
Marijn Storm
mstorm@mofo.com
32 23407364
32-2-347-1824
Jasmine Arooni
jarooni@mofo.com
(202) 887-0763
(202) 572-6759

About Morrison Foerster

We are Morrison Foerster — a global firm of exceptional credentials. Our clients include some of the largest financial institutions, investment banks, and Fortune 100, technology, and life sciences companies. The Financial Times has named us to its list of most innovative law firms in North America every year that it has published its Innovative Lawyers Reports in the region,and Chambers Asia-Pacific has named us the Japan International Firm of the Year for the sixth year in a row. Our lawyers are committed to achieving innovative and business-minded results for our clients, while preserving the differences that make us stronger.

Because of the generality of this update, the information provided herein may not be applicable in all situations and should not be acted upon without specific legal advice based on particular situations. Prior results do not guarantee a similar outcome.

access. In addition to being resilient, secure AI Systems can avoid, protect against, respond to, and recover from attacks.

- **Accountable and Transparent:** Transparency is measured by the extent to which information about an AI System and its outputs is available to individuals interacting with the system. Transparency contributes to the accountability in AI systems, and according to the RMF, accountability is the basis of trustworthiness.

- **Explainable and Interpretable:** Explainability refers to a representation of the mechanisms underlying an AI System's operation. Interpretability refers to the meaning of an AI System's outputs in the context of its designed functional purposes. Explainability and interpretability assist both AI System users and those operating or overseeing an AI System to better understand the functionality and trustworthiness of the AI System and its outputs.

- **Privacy-Enhanced:** Anonymity and confidentiality should guide choices for the design, development, and deployment of AI Systems. Privacy-enhanced AI Systems may utilize data minimizing methods such as de-identification and aggregation.

- **Fair with Harmful Bias Managed:** AI Systems should address issues such as harmful bias and discrimination. The RMF identifies and describes three major categories of AI bias to consider in the management of AI Systems: systemic, computational and statistical, and human-cognitive bias.

The RMF highlights that the characteristics of trustworthy AI Systems should be balanced and applied based on the context of use. Ignoring the characteristics can increase the probability and magnitude of negative consequences throughout the lifecycle of an AI System.

NIST also published an **AI Risk Management Framework Playbook** as a companion resource to the RMF. The Playbook provides suggestions and guidance on how to use the RMF, as well as suggested actions, references, and documentation to help readers achieve the outcomes outlined in the RMF. NIST has solicited public feedback and aims to release a revised Playbook in Spring 2023.