PUBLIC SUBMISSION

As of: 8/19/21 10:45 AM
Received: August 19, 2021
Status: Pending_Post
Tracking No. ksi-zprx-e6z8
Comments Due: August 19, 2021
Submission Type: Web

Docket: NIST-2021-0004
Artificial Intelligence Risk Management Framework

Comment On: NIST-2021-0004-0001
Artificial Intelligence Risk Management Framework

Document: NIST-2021-0004-DRAFT-0023
Comment on FR Doc # 2021-16176

Submitter Information

Email:
Government Agency Type: Foreign
Government Agency: Infocomm Media Development Authority

General Comment

Please see attached document.

Attachments

AI RMF RFI Response V2.2

All comments will be made public as-is, with no edits or redactions. Please be careful to not include confidential business or personal information, otherwise sensitive or protected information, or any information you do not wish to be posted.

Comment Template for Responses to NIST Artifical Intelligence Risk Management

Submit comments by August 19, 2021:

General RFI Topics (Use as many lines as you like)	**Response #**	**Responding organization**	**Responder's name**	**Paper Section (if applicable)**	**Response/Comment (Include rationale)**	**Suggested change**
Responses to Specific Request for information (pages 11,12, 13 and 14 of 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;	1	Infocomm Media Development Authority	-	-	The lack of common metrics that can be used in different AI trust scenarios to measure the different principles in trustworthiness is a key challenge. For example, the safety metrics and risk profile of an AI system used in self-driving vehicles will be different compared to that of a facial recognition system. There is a lack of easy-to-understand assessment tools which AI actors can use to evaluate the trustworthiness of the AI systems. Current state-of-the-art tools have defined their own metrics and is often vague and open to different interpretations. Across the variety of AI use cases, it is a challenge for different AI actors to come to a consistent interpretation of the raw output of tools. There are lack of tools in many classes of AI such as graph neural networks or multiclass machine learning. For instance, it will be a challenge to assess fairness in unsupervised learning and deep learning as most fairness assessment tools support some tasks (e.g. classification) in supervised learning. It is also a challenge to identify a common framework for all stakeholders to agree on the AI-related risks that have to be managed. For example, data scientists may be preoccupied with performance over fairer outcomes, while compliance teams may be preoccupied with risk and liability. Lastly, there is a lack of mitigation measures that can help the AI actors to respond to AI-related risks. For instance, there is no guideline that the data scientists can use to address fairness problem when they identify bias in the datasets that might cause harmful outcome.	-

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;	1	Infocomm Media Development Authority	-	-	We define trustworthiness based on two guiding principles. Firstly, the AI used in the decision-making must be explainable, transparent and fair. Secondly, the AI used should be human-centric with user well-being and safety being primary considerations in designing and deploying the AI solution. In addition, we provide guidance in four key areas that help promote responsible use of AI, in a manner that is algorithm, technology, and business agnostic. The four key areas are: internal governance structures and measures, risk assessment to determine the level of human involvement in AI-augmented decision-making, operations management, and stakeholder interaction and communication to concretely operationalise AI ethics principles. These help organizations adopting the framework set up effective structures, determining acceptable risks, establish good data governance practices, and mitigate bias in data and AI models.	-
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;	1	Infocomm Media Development Authority	-		We think that the following 10 attributes are important principles that must be considered in the Framework. - Transparent (Information about AI system is available) - Fairness (Ensure that there is no unintentional discrimination in the results produce by the AI system) - Accountability (Making sure that AI actor is held responsible for the decision) - **Explainability** (Ability to understand and interpret what AI system is doing) - **Safety** (AI systems must be reliable and wil not cause harm) - **Security** (AI systems must be secured from cyber attacks) - **Data Governance** (Training data must be managed properly and the quality must be ensured) - **Repeatability** (For purpose of verification & validation of the results) - **Reproducibility** (For purpose of verification & validation of the rseults by independent party) - **Robustness** (Ability to handle unexpected input/adversarial attacks) AI systems should be explainable and transparent so the users of the system can trust the predicted results. AI systems that are able to repeat and achieve consistent outputs over multiple runs demonstrate the reliability and stability of these systems. Repeatable results are also a key complement to explainable AI. Being able to reproduce the same results by a third-party team helps to verify claims about the AI systems, allowing the systems to be accredited or certified to perform as claimed, thereby reinforcing trust by users. AI systems should also be robust to handle unexpected input to produce correct output. Such unexpected input may or may not be due to adversarial attacks (e.g., falling trees covering part of traffic signs but the autonomous vehicle would still be able to recognise the traffic signs correctly). This would help to ensure the safety of the AI systems such that it will not cause harm. AI systems must also be secured from cyberattacks to ensure that they remain safe and accurate for the users of the system.	-
4. The extent to which AI risks are incorporated into different organizations' overarching enterprise risk management – including, but not limited to, the management of risks related to cybersecurity, privacy, and safety;	1	Infocomm Media Development Authority	-		We believe that AI risks should be contextualized with respect to an overarching risk management framework. In most cases, AI-related risks (e.g. data privacy, security of the system) should be aligned to the organization's internal policies and compliance. This ensures the practices and standards applied in other areas are also consistently applied on the AI systems. It will improve trust in these systems and simplify stakeholder communications on managing AI-related risks as an extension of existing policies.	-

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;	1	Infocomm Media Development Authority	-		- Singapore's Model AI Governance Framework - ISO/IEC TR 24028:2020: Overview of trustworthiness in AI	-
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;	1	Infocomm Media Development Authority	-	-	Currently, there is no regulatory requirements to mandate the use of specific standards or best practices. Organisations are encouraged to adopt the guidelines provided by Singapore Government voluntarily. The Singapore Government provides guidelines such Model AI Governance Framework and Self-Assessment Guide for Organisations (ISAGO) for organizations on key considerations and measures to be implemented in designing and deploying the AI solution.	-
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;	1	Infocomm Media Development Authority	-	-	Singapore's Model AI Governance Framework	-
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.	1	Infocomm Media Development Authority	-	-	Organizations should weigh in all the possible damages that could happen to the organization when such issues arise. Organizations might develop poor reputation if they are found to use or develop unfair AI systems. This could erode the customer base and hinder future businesses that are done by the organizations. Whereas for AI systems that are properly designed and tested to be inclusive will increase the value of the organizations. The developers of the AI systems should be made aware of the negative impact on individuals, groups, and society when it comes to inclusiveness in AI development. AI developers should use available tools to properly test the datasets and machine learning models. This can help to highlight any potential bias issues that could arise in the model development phase. Apart from that, there should be a proper process in ensuring accountability in the decisions that are made by the AI developers. In general, organizations should take reasonable efforts to address multivariate sources of bias in AI. Specifically, it is critical that datasets used for AI model training are suited for purpose and do not propagate inherent flaws in data to effectively manage the risks of inaccuracy or bias. Organizations should institute processes to understand how datasets could be inherently biased and develop a strategy to detect, understand and mitigate these issues. For instnace, organizations may have to develop mechanisms around accurate data tagging, periodic review, and validating sources of datasets. Where the dataset permits, organizations may apply different datasets for training, testing and validation to minimize negative impact.	-
9. The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes");	1	Infocomm Media Development Authority	-		Highly appropriate	-

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	1	Infocomm Media Development Authority	-	-	Singapore's Model AI Governance Framework highlights the importance of conducting risk assessment prior to the design of an AI system. This framework suggests risk dimensions to consider, namely, severity of impact and possibility of impact on the individuals, taking into consideration context of use. Based on the risk assessment, AI developers can build in risk mitigating measures including the level of human involvement in AI-augmented decision making and the need to shut down AI system gracefully. The Framework can also be effectively structured in such a way that it provides AI actors with guidelines to test the principles against the AI systems apart from integrating AI risk processes with organizational processes. The risks could be managed but it will be much more effective to include testings to validate and verify the claims for developing AI products and services.	-
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.	1	Infocomm Media Development Authority	-	-	Today, it is difficult to recruit the suitable candidate to run AI-related functions within the organizations as most of the AI-related risks are siloed among the AI actors (e.g. business groups vs technical groups). The Framework could lay out the different aspects of AI-related risks (both non-technical risks and technical risks) that can occur at various stages in the lifecycle of the AI systems. Each risk can be elaborated with guidelines that contain detail on how to assess, test and respond to. This will set a clear indication to the DevSecOps, AI and audit community the skillsets that are required to perform AI-related functions within the organization. It will also be useful to reference other standards, methodologies or frameworks that are relevant or complementary to AI design, development, and deployment. These frameworks collectively improve upon our appreciation of AI risks, and will chart essential skills for workforce learning and development, similar to other frameworks of knowledge used in the technology sector.	-
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.	1	Infocomm Media Development Authority	-	-	The Framework should include detailed guidelines on how organizations can adapt existing or set up internal governance structures and risk mitigation measures. However, guidelines need to retain a degree of flexibiltiy for it to be adapted to different business sectors and risk profiles. With reference to Singapore's Model AI Governance Framework, this guides organizations to recognise and document AI-related risks in their internal policies and assign clear roles and responsibilities for the design and deployment of AI solutions.	-