Response to the Request for Comments on the NIST AI RMF Playbook

Intel Corporation appreciates the opportunity to provide comments to the National Institute of Standards and Technology ("NIST") on AI RMF (Risk Management Framework) and playbook. We appreciate NIST's work in defining a risk-based approach to Artificial Intelligence (AI) analysis.

Intel's mission is to engineer solutions to address society's greatest challenges, providing our customers with reliable, cloud-to-edge computing, inspired by Moore's Law. Intel designs and manufactures advanced integrated digital technology platforms that power the connected world.

Intel is committed to advancing AI technology responsibly and contributing to the development of principles, international standards, best practices methods, tools, and solutions to enable a responsible, inclusive and sustainable future. We commend NIST's efforts in developing an AI Risk Management Framework to support stakeholders of AI systems in managing risks across the AI lifecycle.

General Considerations

Similarly to other broad horizontal spaces, Artificial Intelligence (AI) covers a range of systems and approaches that form a continuum. In this note, we would like to highlight one of the aspects of the AI systems: the distinctions between human facing and non human-facing AI systems. Human facing systems focus on human-centric problems, are directed at human users, and typically rely on sensitive data such as Personally Identifiable Information (PII). Non human-facing systems typically produce analytics for other AI systems, focus on addressing other types of problems (e.g., terrain analysis or weather metrics) and typically use data with no or minimal PII components. AI systems can fall in between these two extremes.

When building frameworks to cover all AI systems, a question which frequently arises is if one framework can cover the whole range of possibilities including the extremes. The answer to this question is yes, if the framework recognizes these differences and provides flexibility based upon different use cases. Most of the evaluation and risk-based frameworks existing today are based on a canonical list of risk criteria and enable the evaluators to determine which criteria

are relevant to a particular system. The ability to cover a wide range of use case is also characteristic of frameworks that include a significant societal aspect, e.g., the NIST Cyber Physical Systems framework[1]. This property is present also in the frameworks listed for comparison in the 2nd draft of the RMF framework (p.17).

While comprehensive frameworks have the ability to cover a wide range of use cases, recognizing human facing and non human facing AI systems, many frameworks cover only human facing systems use cases. Here, we would like to offer an example of using one common framework to cover human-facing and non human-facing AI systems. Emerging approaches already explicitly include a wide range of systems, e.g., recent work at Berkeley CLTC[2]. The approach used by CLTC creates a continuum between human-facing and non human-facing systems, based on several parameters.

Below we provide a simplified illustration of the applicability of criteria in a common framework similar to draft NIST AI RMF.

Consider an AI system that collects information from weather sensors in a certain geographic area, aggregates the results, and provides primary analysis according to a predefined AI model. The analytics are fed to a regional weather forecast system, while raw data are also submitted to the regional system, following pre-processing. The data in question, its origin and destination as well as AI analytics include no privacy considerations. Therefore, evaluation for privacy risks can be omitted in this case. However, evaluation of data quality and data completeness are important, putting considerable weight on the need for accuracy. For example, if the data are collected only in sunny areas or only between 6am and 7am or if some sensors are faulty, it will be difficult to make evidence-based conclusions, elevating the risk of insufficient accuracy.

In Table 1, we compare the set of criteria relevant to the non human facing weather metrics system described above with the set of criteria to be used for an average human facing systems. Here and further in the document, we use the risk criteria of a generic human facing system as a gauge to compare with the use case under evaluation. We use this approach since all the examples provided so far during the draft NIST AI RMF discussions focus on human facing systems.

Table 1. Evaluating some risks of machine to machine focused local weather monitoring system

Criterion	**Human Facing**	**Weather system (non human facing)**
Accuracy	Yes	Yes
Reliability	Yes	Yes
Robustness	Yes	Yes
Resilience	Yes	Yes
Security	Yes	Yes

[1] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-201.pdf

[2] Properties of Trustworthiness for Artificial Intelligence: A Framework for Algorithmic Governance & Quality, Intel and the UC Berkeley Center for Long-Term Cybersecurity [Forthcoming]"

Explainability	Yes	Evaluated differently
Interpretability	Yes	Yes
Privacy	Yes	No
Safe/Safety	Yes if applicable	Yes if applicable
Fairness	Yes	At development stage, with regard to the development/maintenance teams
Accountability	Yes	Yes
Transparency	Yes	Yes

Human facing and non human facing AI applications

As described above, a continuum exists, with the extremes of fully human facing or fully non human facing AI systems, with many systems falling in between these two extremes. As the example above illustrates, all systems on this continuum can be evaluated using criteria developed by NIST AI RMF and using the related playbook. In the NIST AI RFM, most criteria apply across the board; however, the relative weight of a criterion may depend on the nature of the use case. For example, an AI system associated with critical parts of our infrastructure such as power plants, where inaccuracies can lead to catastrophic failures, needs to attach greater weight to accuracy, security, robustness, resilience, and security, while an AI system predicting selection of colors for clothing can also be evaluated along the same technical criteria, but with less weight on criteria that are not critical.

In some criteria, there may be a greater distinction between human facing and non human facing systems than in others that always apply across the board. As illustrated earlier, privacy may not be relevant for systems that are machine-to-machine oriented with no personal data in the data sets. Explainability may take different forms for computationally intensive systems with a lot of data that are not user facing, whereas this criterion has a much greater impact for human facing systems. Safety may be a crucial criterion for some AI systems regardless of their human facing or non human facing nature.

Finally, the principles developed by various AI frameworks apply in all cases, but may be used differently in different use case, e.g., requirements for transparency for confidential or critical infrastructure systems would need to be evaluated and implemented differently compared with consumer targeted systems.

The NIST AI RMF 2nd draft provides extremely useful characteristics of Trustworthy AI in Section 4.0. We summarize the reasoning for broad applicability of the NIST AI RMF framework and playbook below. It is obvious that risks in different systems can be measures using uniform criteria, but different system requirements. For example, accuracy of matching resumes to job descriptions and accuracy of a local weather station metrics relies on different parameters. But these differences don't create an obstacle for the uniform evaluation of an AI system.

Table 2. Example risk evaluation criteria and their typical applicability across the human facing to non human facing continuum

Criterion	Human Facing	Mixed	Non human facing
Accuracy	Yes	Yes	Yes
Reliability	Yes	Yes	Yes
Robustness	Yes	Yes	Yes
Resilience	Yes	Yes	Yes
Security	Yes	Yes	Yes
Explainability	Yes	Frequently	Not always, may have different requirements
Interpretability	Yes	Yes	Yes
Privacy	Yes	Frequently	Not always, depending on the nature of the system
Safe/safety	Yes, if applicable to the use case	Yes, if applicable	Yes, if applicable
Fairness	Yes	If applicable	Almost never
Accountability	Yes	Yes	Yes
Transparency	Yes	Frequently	Frequently

Use cases for illustration

We provide additional use cases for illustration of the use of the NIST AI RMF Framework below.

Procurement, including but not limited to the development of sourcing strategies, select, and manage supplies, contract management: Develops sourcing strategies by using organizational historical data to build spending profiles and procurement plans based on demand forecast, inventory levels, and leveraging internal and external data; categorize spend categories to help set strategy and monitor performance.

Table 3. Evaluate a mixed (human facing/non human facing use case) using the NIST AI RMF approach (Human Facing column is provided for comparison)

Criterion	**Typical Human Facing**	**Use case: Procurement system (mixed)**
Accuracy	Yes	Yes
Reliability	Yes	Yes
Robustness	Yes	Yes
Resilience	Yes	Yes

Security	Yes	Yes
Explainability	Yes	Yes
Interpretability	Yes	Yes
Privacy	Yes	Yes
Safe/Safety	Yes, if applicable	Partially
Fairness	Yes	Yes
Accountability	Yes	Yes
Transparency	Yes	Yes, depending on the level of confidentiality

Supply chain function including but not limited to managing demand for product and services, manage materials, operate warehousing, manage distribution: AI is deployed to develop baseline forecasts for product and services using historical data and external factors; build the consensus forecast; identifies drivers, correlations, and trends and conduct automated forecasting including best-fit algorithm selection; Identify critical material and supplier capacity; determine and monitor finished goods inventory requirements at the destination using historical data and taking into consideration external market data.

Table 4. Evaluate a non human facing use case using criteria similar to the NIST AI RMF approach. The Human Facing column is provided for comparison purposes.

Criterion	**Human Facing**	**Use Case: Supply chain system (non human facing)**
Accuracy	Yes	Yes
Reliability	Yes	Yes
Robustness	Yes	Yes
Resilience	Yes	Yes
Security	Yes	Yes
Explainability	Yes	Yes
Interpretability	Yes	Yes
Privacy	Yes	Limited
Safe/Safety	Yes, if applicable	Yes, If applicable
Fairness	Yes	Yes, during the development and maintenance
Accountability	Yes	Yes
Transparency	Yes	Yes, depending on the level of confidentiality required in the system

Terrain analysis for use in GIS (Geographic Information Systems) and other logistics-based applications. Terrain characteristics are collected via various options, from on road data collection to satellite image collection. The data are merged and insights from analytics fed to

various systems, including GIS. Data are not meant for direct human use, but datasets may include elements of PII (e.g., images of housing or private roads).

Table 5. Evaluate a non human facing use case with approaches similar to the NIST AI RMF approach. The Human Facing information is provided for comparison purposes.

Criterion	**Human Facing**	**Terrain analysis system (non- human facing)**
Accuracy	Yes	Yes
Reliability	Yes	Yes
Robustness	Yes	Yes
Resilience	Yes	Yes
Security	Yes	Yes
Explainability	Yes	Yes
Interpretability	Yes	Yes, according to the requirements
Privacy	Yes	Yes, if data contain PII
Safe/Safety	Yes if applicable	Yes
Fairness	Yes	Yes, during the development and maintenance
Accountability	Yes	Yes
Transparency	Yes	Yes, depending on the level of confidentiality required in the system

In conclusion

The universe of AI systems is very broad and includes a continuum from human-facing to non-human facing systems. While the focus on risk management approaches has been primarily based on human facing systems, differentiating between human-facing and non-human facing applications will enrich frameworks such as NIST AI RMF version 2, by allowing the developers and users to incorporate soft and hard metrics and weigh the impact of criteria for use cases while preserving uniformity of the approach. Including non-human facing systems as examples permits the stakeholders to evaluate a broad range of AI applications using a unified framework approach while recognizing that the potential risks can be very different.

We commend NIST for undertaking such an enormous task in developing a complex and flexible Risk Management Framework and Playbook for AI. If you have any questions on this note or are interested in further discussion on this subject, please contact Claire Vishik ( ) and Grace Wei ( ).