KAISER PERMANENTE®

Kaiser Foundation Health Plan
Program Offices

April 29, 2022

RE: *NIST AI Risk Management Framework: Initial Draft*

Submitted via email to: **AIframework@nist.gov**

Kaiser Permanente (KP) appreciates the opportunity to offer feedback on the above-captioned request for comments (RFC).[1] The Kaiser Permanente Medical Care Program is the largest private integrated health care delivery system in the U.S., delivering health care to over 12 million members in eight states and the District of Columbia[2] and is committed to providing the highest quality health care.

The use of Artificial Intelligence (AI) is rapidly expanding, including in many clinical, administrative and research functions in healthcare, and its growth is expected to continue for the foreseeable future. NIST has tackled many critical aspects of AI including managing bias, explanation, and trustworthiness. We appreciate NIST's current efforts to develop an AI Risk Management Framework (AI RMF) to guide institutions and for AI development and use in systems. We offer the following in response to questions posed:

1. **Whether the AI RMF appropriately covers and addresses AI risks, including with the right level of specificity for various use cases?**

The AI RMF appropriately covers and addresses AI risks and includes the right level of specificity for a variety of use cases, however, the following areas would benefit from additional content and detail:

- Algorithmic Impact Assessment (AIA): The AIA is a fundamental component necessary for risk understanding and management, and it enhances the identification and management of AI risk. We recommend that the AI RMF incorporate an AIA.
- Coverage regarding inputs: The AI RMF should include coverage regarding inputs, such as threats and their motivation, underlying system vulnerabilities, exploitability of vulnerabilities and algorithms.
- Vulnerability: The AI RMF should include a discussion about assessing the likelihood for a vulnerability to impact an AI system because vulnerabilities differ in likelihood. Risks related to system configuration management, real-time monitoring, model brittleness, data drift, and AI algorithm degradation over time should be discussed, including operational perspectives.
- Updated AI lifecycle: We recommend tying the updated AI lifecycle, and definitions of bias, to the three broad areas NIST identified that present challenges for addressing AI bias. The AI RMF would benefit from this structural change and could be brought to life with more specificity, for example through one use case taken end to end through the model lifecycle. We also recommend tying the

[1] https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

[2] Kaiser Permanente comprises Kaiser Foundation Health Plan, Inc. and its health plan subsidiaries outside California and Hawaii; the not-for-profit Kaiser Foundation Hospitals, which operates 39 hospitals and over 720 other clinical facilities; and the Permanente Medical Groups, self-governed physician group practices that exclusively contract with Kaiser Foundation Health Plan and its health plan subsidiaries to meet the health needs of Kaiser Permanente's members.

updated AI lifecycle to the CRISP-ML[3] flow and interweaving a series of checkpoints along the lifecycle for stewards to mitigate and reduce bias (e.g., during data analysis, production).

- Use case examples: We recommend using different types of use cases as examples in the AI RMF, rather than solely machine learning (e.g., natural language processing, Blackbox).
- AI Risk Management Fact Sheet: NIST should publish an AI Risk Management Fact Sheet similar conceptually to the one NIST created for C-SCRM (May 2021).

2. Whether the AI RMF is flexible enough to serve as a continuing resource considering evolving technology and standards landscape?

The AI RMF is an excellent basis for future planning and AI and ML model risk management. The document is very flexible. It would be enhanced by providing policy guidance and documentation templates that organizations could leverage in an appendix. NIST should also consider carving out Governance of AI and publishing a specific document on this topic.

3. Whether the AI RMF enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks?

The AI RMF sufficiently enables decisions about how an organization can increase understanding of, communication about, and efforts to manage AI risks. However, an additional section, appendix or separate document that includes evidence and real-world examples of the impact of bias or perceptions of unfairness would be beneficial.

Communication to senior management is a critical component of AI risk management in an organization. We recommend that the AI RMF include a figure similar in nature to Figure 3 in NISTIR 8286C (Draft) to support management education and discussions (e.g. Identify → Measure → Mitigate → Risk). We also recommend that the AI RMF highlight the capabilities for actionable decisions embedded within a defined process and include reporting that identifies deviations from governance, policy, risk tolerance and standards.
Examples of governance and stakeholder engagement models would also be helpful to better understand how AI risks should be communicated and socialized throughout the organization.

4. Whether the functions, categories, and subcategories are complete, appropriate, and clearly stated?

Identification of the three dominant categories of AI bias is robust, however, it is unclear how these biases and sub-biases contribute into the overall AI RMF. We recommend additional language for AI model verification and validation as a specific risk measure to be included in Table 2, ID #3.

5. Whether the AI RMF is in alignment with or leverages other frameworks and standards such as those developed or being developed by IEEE or ISO/IEC SC42?

The AI RMF is in alignment with other related frameworks and standards. For example, IEEE has invested in and published ethical design and principles of AI, supported by NIST, focusing on a broad suite of technical issues associated with AI. We see no conflicts with the work of IEEE or ISO/IEC SC42.

6. Whether the AI RMF is in alignment with existing practices, and broader risk management practices?

[3] Cross-Industry Standard Process for Machine Learning, https://ml-ops.org/content/crisp-ml

The AI RMF aligns with traditional risk management existing practices, however, the key areas of organizational policies and standards need additional exploration and development in the context of AI. AI is a rapidly evolving field with ever-increasing risk due to many advancements, and the AI RMF may require frequent revisions until the field stabilizes.

7. What might be missing from the AI RMF?

We recommend NIST consider the following issues for inclusion in the AI RMF:

- Knowledge Brittleness: The concept of knowledge brittleness, and related boundary condition design issues for models was highlighted in the most recent NIST AI RMF Workshop. Quite simply, when, and how, does the model know and react -- to what it doesn't know? This notion is especially critical for life-safety systems such as those in health care and we recommend NIST consider how to address this concept in the AI RMF.
- AI Risk Management with Vendor/3rd party/ Suppliers: AI risk associated with vendors, third parties, and the software supply chain is an essential component of AI risk management and we recommend that the AI RMF address this topic in more detail.
- Transparency: The AI RMF should include a more thorough discussion of risk associated with Transparency, including the traceability/provenance and auditability of data across the lifecycle spectrum from training and testing to real-world operations and updates.
- AI lifecycle: The AI RMF should include a specific discussion intended for data scientists and model developers on the tangible steps they can take during the updated AI lifecycle to mitigate and document AI risks that arise from the three dominant categories of bias. Additional focus on technical frameworks, code snippets, and tools highlights would be beneficial.
- Sustainability: Responsible AI considers the sustainability aspects of the AI systems and methodologies to measure and assess computational resources and cost of the AI application's training, development, monitoring, and maintenance. We recommend that the AI RMF include a discussion regarding sustainability and how to incorporate those results into the risk governance process.

8. Whether the soon to be published draft companion document citing AI risk management practices is useful as a complementary resource and what practices or standards should be added?

A companion document citing AI risk management practices as a complementary resource for AI risk management practices could be very useful. This should contain use cases targeting at least four or five domains such as financial services, healthcare, pharmaceuticals, energy, or other critical national infrastructure domains.

9. Any additional comments?

We encourage NIST to reconsider whether the AI RMF properly:

- Addresses risk quantification associated with the relative strength of explanation or degree of model inscrutability. The probability, frequency, and loss magnitude should be considered when seeking to quantify risk. Risk quantification also should include the costs of data acquisition or loss, costs to maintain and rebuild, costs to comply with regulations, and reputational costs (societal trust).

- Describes concepts of quantitative measurement of AI Risks
- Discusses the concept of Trustworthiness in the context of AI in healthcare [technical, socio-technical or principles]. KP recommends incorporating the notion of natural language processing or NLP as an additive component for risk explanation to aid in trustworthiness and end user acceptance.
- Describes instantiations of an AI-Risk Management Profile in the healthcare context. A strong emphasis on health care use cases targeting both clinical and non-clinical would be beneficial in describing the health care instantiations of AI Risk management.

We also offer the following proposed technical revisions and amendments:

- Section 6.3 Manage Table 3, ID 3: Include Business Continuity Disaster Recovery or High Availability Disaster Recovery to this statement, "Plans related to post deployment monitoring of the systems are implemented, including mechanisms for user feedback, appeal and override, decommissioning, incident response, and change management".
- Section 6 AI RMF Core line 12-14 and section 6.4 Govern: Consolidate these sections to avoid duplication.
- Section 6.4: Remove "Governance should address supply chains, including third-party software or hardware systems and data, as well as internally developed AI systems." Supply chain security should be addressed by Vendor Risk Management, a sub-component of Enterprise Risk Management. AI Risk Management Function is also a sub-component of Enterprise Risk Management.
- Section 6.4: Move "Govern" to the top. This section sets the strategy, policy, standards such as AI security standard, which AI RMF would use as input. Replace the word "practicable" with feasible or achievable.
- Section 6.4 Table 4 ID 4: Clarify the statement "Organizational practices are in place to ensure that teams actively challenge and question steps in the design and development of AI systems to minimize harmful impacts." Is this statement trying to state the practice of Threat Modeling? We recommend as part of any assessment; criteria should be based on industry best practices and reference architecture.

* * *

We applaud NIST for this valuable and thoughtful work. Please feel free to contact Jamie Ferguson or Megan Lane with any questions or concerns.

Sincerely,

Jamie Ferguson
Vice President, Health IT Strategy and Policy
Kaiser Foundation Health Plan, Inc.