KAISER PERMANENTE®

Kaiser Foundation Health Plan
Program Offices

September 29, 2022

RE: *AI Risk Management Framework: Second Draft for public comment*

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.

As a health care organization, KP is committed to promoting innovative technologies that have the potential to transform the delivery of health care and improve patient care. We support NIST's continued work to develop a framework to address risks in the design, development, use and evaluation of AI products, services, and systems.

We offer the following comments:

General Comments

AI systems can amplify or exacerbate inequitable outcomes in a variety of settings, including health care. For this reason, it is important to work rapidly towards more specific recommendations on approaches, practices, standards, metrics, and methods to achieve measurable trustworthiness and responsibility of cognitive systems.

We recommend focusing efforts to tailor the AI RMF towards proving trustworthiness and responsibility as part of a cognitive systems lifecycle. End users should have a mechanism to assess predictions and decision proposals that result from AI cognitive functions, much like we do in real world scenarios when assessing experts and whether to rely on their advice. Lifecyle oversight could be expanded and improved via open participatory governance of cognitive systems combined with input from consumer advocacy groups associated with use of the system.

Overview

Trustworthy and Responsible AI

AI systems can operate with high levels of autonomy with substantial impact potential to individuals, communities, and society. We recommend that NIST explore a deeper risk analysis model to expand on key differences between AI and non-cognitive software systems. For example, as cycles between training and inference shorten, cognitive systems become non-deterministic and system errors become

[1] https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.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.

non-reproducible. We also recommend that the risk management framework for cognitive systems incorporate or reference practices to mitigate risk used in other fields such as those used in clinical trials, pharmacology, and other evidence-based methodologies and scientific research.

Purpose of the AI RMF
We agree that the purpose of the AI RMF is to cultivate trust by understanding and managing the risks of AI systems, however, we recommend that the purpose of the RMF also include informing users and impacted groups of how risks associated with cognitive function outputs have been managed. We recommend referencing the dynamic and ever-evolving nature of risk management in this section as well to illustrate how our understanding of fairness, bias, ethics, etc. in AI is constantly evolving.

Audience
We believe that the large impact of data and its accuracy and appropriateness in cognitive systems warrants stronger focus on the "data and input" step in the AI lifecycle. We recommend developing ways to track provenance in the framework such as data fingerprinting or blockchain-based time sequencing to trace the information pedigree of a cognitive system's output.

We also recommend that the user stakeholder group differentiate between the following subgroups to assure trustworthiness:

- People directly using applications with cognitive functions,
- Advisors to other people, and
- Suppliers/consumers in supply chains for cognitive function where AI-systems may interact with each other.

Framing Risk
We recommend that NIST include references to Green AI[3] in this section and encourage sustainable development and evaluation of AI when possible.

AI Risks and Trustworthiness
Intentions must be monitored more closely when the degree of autonomy of an AI system is higher, and the framework should reflect this. Fully autonomous decision making where intent might skew malicious requires enhanced attention and monitoring compared to human-in-the-loop AI where scope/intent of the business use case is restrained systematically. We also recognize that the goals of transparency, explainability, and interpretability can be at odds with or even inversely correlated to the goal of building a highly accurate, private, secure, or fair system, depending on the use case, but all these goals are beneficial. To help achieve all these goals together we recommend that NIST carefully consider and calibrate these priorities.

Privacy requirements, rules, and regulations vary substantially for different industries and domains. In the health care sector, there are additional privacy requirements arising from federal law, such as the Health Insurance Portability and Accountability Act (HIPAA), guidance from federal agencies such as

[3] The term as used here refers to AI research that yields novel results while considering the computational cost, encouraging a reduction in resources spent.

the Federal Trade Commission (FTC) and Office for Civil Rights (OCR), state law, and international regulations such as the European General Data Protection Regulation (GDPR) that should be respected during AI development. We recommend that NIST take these considerations into account in the Privacy section.

Effectiveness of the AI RMF

Effectiveness is one of the most important considerations for an AI-RMF. It is crucial to develop and implement standards, measures, and tools that help organizations measure and understand AI risk consistently across organizations and sectors. The use of standardized criteria and standardized data definitions will help to ensure consistent, reliable measurement of AI risk, and speed user community acceptance. Such metrics will support trustworthiness of model outcomes.

AI RMF CORE

The four functions proposed (Map, Measure, Manage, Govern) are single-dimensional elements that should be better aligned with the overarching NIST RMF[4] that includes seven functions: Prepare, Categorize, Select, Implement, Assess, Authorize, Monitor (decay, detect, respond, recover).

We also recommend that the Governance section include a broader focus to cover areas of financial, operations in addition to supply chains.

* * *

We applaud NIST for this valuable and thoughtful work and look forward to the release of a more thoroughly developed framework. 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.

[4] https://csrc.nist.gov/projects/risk-management/about-rmf