COMMENTS OF THE
ACM US TECHNOLOGY POLICY COMMITTEE
ON MARCH 17, 2022 INITIAL DRAFT OF
NIST ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK

Introduction

The Association for Computing Machinery (ACM), with more than 50,000 U.S. members and approximately 100,000 worldwide, is the world's largest educational and scientific computing society. ACM's US Technology Policy Committee (USTPC), currently comprising more than 160 members, serves as the focal point for ACM's interaction with all branches of the US government, the computing community, and the public on policy matters related to information technology. USTPC responded in August 2021 to NIST's Request for Information on its Artificial Intelligence Risk Management Framework (Docket Number 210726-0151)[1] and is pleased to again contribute to the evolution of this important effort. To that end, we respectfully submit the following over-arching suggestions on the Initial Draft Framework released on March 17, 2022, supplemented by "line level" comments, as follows:[2]

General Analysis and Recommendations

1. **Risk Ranking --** USTPC notes that not all risks are created equal and, therefore, not all risks should be met with identical responses. We thus urge NIST to establish a hierarchy of enumerated risk categories. Under such a system, particularly profound and significant risks placed in a top "tier" would demand the highest levels of system integrity and the most aggressive and active risk management. For example, risks to human life presumably would be ranked highest while, for example, applications affecting regulated spheres such as hiring, credit, housing, or allocation of public resources might occupy a second tier, etc. The scope of a system's impact also should be considered; a system with a million customers or users thus would require more safeguards than one with a thousand.

[1] *See*, Comments on National Institute of Standards and Technology RFI re Artificial Intelligence Risk Management Framework, ACM US Technology Policy Committee (August 19, 2021). In December 2021, a number of USTPC's expert members also jointly submitted personal comments on the agency's AI Risk Management Framework Concept Paper.

[2] Principal authors of this submission for USTPC were its AI & Algorithms Subcommittee Co-Chair Jeanna Matthews, Past USTPC Chair Stuart Shapiro, and Committee member Ricardo Baeza-Yates.

ACM U.S. Technology Policy Committee
1701 Pennsylvania Ave NW, Suite 200
Washington, DC 20006
+1 202.580.6555
acmpo@acm.org
www.acm.org/public-policy/ustpc

2. **Guideline Specificity --** We encourage NIST both to set clear guidelines to help organizations identify appropriate specific integrity levels for their proposed systems and to then map such rankings to a recommended set of specific risk management activities for that integrity level. USTPC further urges NIST in this regard to specifically recommend that such risk management activities include the development of processes for:

    - system verification and validation;
    - supporting auditing decisions in cases where harm is suspected;
    - maintaining data provenance;
    - enabling questioning by, and redress for, individuals and groups that are adversely affected by algorithmically informed decisions;
    - independent verification and validation of all systems in the highest risk tiers; and
    - deciding whether an AI system should be built at all based on legal, ethical, and scientific risk assessments.

    We also encourage the robust exploration and delineation of risk management activities designed explicitly to protect individuals or groups that may directly experience potential harm or inequities as the result of a particular AI system's operation. Such activities profitably could include their inclusion and involvement throughout a system's design, implementation, testing, deployment, reevaluation, and maintenance. We especially recommend algorithmic *impact* assessments. While such assessments and algorithmic risk assessments are similar in flavor, impact assessments focus on effects on individuals and society (as well as potential harms) more than on the management of risks as perceived by those developing and deploying the AI and automated decision making (ADM) systems.

3. **Context Awareness --** Decision makers often develop or purchase AI or ADM systems to increase their own decision-making efficiency. In addition, managing the risks that impact individuals, classes of individuals, or society as a whole can be costly and at odds with the desired speed of development/ deployment and decision-making efficiency. USTPC thus suggests that market forces may often be insufficient to adequately manage the aggregate risks to society and risk of harm to individuals posed by algorithmically driven systems. As with other complex systems, therefore -- like food or pharmaceutical safety where it is difficult for individuals to truly inspect the risks from the perspective of users and consumers -- government action may be necessary to establish and potentially to enforce standards to protect system users.

USTPC recognizes that NIST does not possess regulatory authority. We believe it would be highly impactful, however, for the Framework to effectively and explicitly list specific recommended risk management activity and system integrity standards linked to defined risk tiers. AI system deployers could be encouraged to clearly argue which risk tier they believe applies to their system and why, and then to disclose which of the recommended conditions for that risk tier have and have not been met.

4. **Definitional Precision --** We note the highlighted comment on page two of the Initial Draft Framework that: "For the purposes of the NIST AI RMF the term artificial intelligence refers to algorithmic processes that learn from data in an automated or semi-automated manner." USTPC encourages NIST to expand that definition to include automated decision-making systems more broadly. The risks, especially from closed box decision-making systems, are similar regardless of whether such systems are learning from data in an automated or semi-automated manner or encoding decision making rules in some other form.

Line-Level Observations

Beyond the foregoing broad comments, USTPC also respectfully submits the attached Appendix with detailed annotations to specific text in the Initial Draft Framework. A number of these comments address issues of structure and terminology, involving in particular the characteristics of AI risk and some of the proposed sub-categories. We also note in a number of instances the importance of explicitly providing for value and ethically based evaluations early in the life-cycle of a system's development and of viewing such evaluations as distinct from risk assessments.

Conclusion

USTPC appreciates this opportunity to again provide input to this important process. For additional information, or to further access the expertise of USTPC and ACM members, please contact Adam Eisgrau, ACM Director of Global Policy & Public Affairs, directly at 202-580-6555 or eisgrau@acm.org.

Appendix

USTPC ANNOTATIONS: AI RISK MANAGEMENT FRAMEWORK INITIAL DRAFT

PAGE	LINE(S)	COMMENTS
8, 9	34, 2	There are two levels of uncertainty in operation: that of the model and that of the risk management methods. We believe it behooves practitioners to maintain awareness of both types.
7	3 - 6	As written, this suggests a chicken and egg scenario wherein a system must be developed and assessed in order to conclude that it should not be developed. It should be made clear that under some circumstances risk management is moot because societal values may militate from the outset against the proposed use case.
7	17 - 21	The act of thinking through risk-related issues in a structured way can prove beneficial even in the absence of defined thresholds.
7	24 - 26	ERM processes may not be architected to deal with the kinds of normative issues potentially raised by AI. They thus might have to be altered accordingly.
8	2 - 3	The asserted inverse relationship between trustworthiness and risk is too reductionist in its framing and sidesteps the distinction between subjective perception and objective measurement.
8	17 - 18	It is not necessarily the case that all technical characteristics would be under the control of system designers and developers, nor that factors under the control of designers and developers are exclusively technical. Reframe more precisely.
10	12	It is noted early in the Framework that AI encompasses more than just ML, but that AI is not necessarily the same as ML. In other instances that follow, however, the terms appear to be used interchangeably. This conflation has the potential to confuse users of the document. We urge that the two terms consistently be appropriately distinguished from one another.
11	11 - 12	The possibility that the model is not operating as expected is not an explain-ability risk; rather, explainability should serve to reveal operational risk.

12	5 - 8	The safety of a socio-technical system inherently has a strong technical component, one which is reflected in the approaches listed. The phrasing used seems to inappropriately discount the technical aspects of safety.
12	10	Bias categories are here presented as if they are all self-explanatory, though they are not. Moreover, computational bias might be more properly thought of and categorized as a technical characteristic.
13	23 - 24	The definition of transparency seems to imply that it might be more properly considered a socio-technical characteristic.
15	13 - 15	This seems to cast the question of appropriateness purely in technocratic terms, excluding the possibility of value or ethically based determinations.
16	2- 3	The preceding category involves assessing, but the data necessary to do so is addressed here. Reorganization or cross-reference may be useful.
Table 1	ID 1	We commend the emphasis on stakeholder engagement in the fourth subcategory and urge NIST also to add it to the first and last subcategories.
Table 1	ID 2	The last subcategory re: context, would seem more aligned with ID 1.
Table 1	ID 4	The second subcategory seems to present the system as a fait accompli, leaving no room for deciding against its development in the first place. Similarly, the phrasing of the last subcategory, with benefits outweighing risks, reads as an assumption rather than an outcome.
Table 3	ID 1	The first subcategory addresses activities that would happen early in the development life cycle, yet it is not addressed until this late stage of the Framework. While we appreciate that risk management activities are intended to happen throughout the lifecycle, this organizational structure of the Core is counterintuitive and potentially confusing.
Table 3	ID 2	Is the objective in the second subcategory to sustain (business) value or to maintain risk posture? We believe the latter is more to the point.
Table 4	ID 1	We recommend that this category also explicitly include ERM integration.
Table 4	ID 4	The wording of the first subcategory conveys the sense that this is optional rather than necessary, while the second subcategory leaves unaddressed the ability to question the appropriateness of a project itself.