Computing Community Consortium (CCC) Response to National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework

September 2021

Brian LaMacchia (Microsoft Research), Daniel Lopresti (Lehigh University and Computing Community Consortium), and Helen Wright (Computing Community Consortium)

Response to Request for Information (RFI) on the NIST Artificial Intelligence Risk Management Framework (AI RMF or Framework):
https://www.federalregister.gov/documents/2021/07/29/2021-16176/artificial-intelligence-risk-management-framework

In 2018-2019, the Computing Community Consortium (CCC) brought together over 100 members of the research community to create *A 20-Year Community Roadmap for Artificial Intelligence Research in the US* (AI Roadmap). We offer our responses to the following points from the RFI drawing from the extensive discussions within the national AI research community that arose while developing the AI Roadmap. We appreciate that NIST recognizes AI and its risks are a moving target and there will be a need to make regular updates to the RMF in the future. We do not believe it is an exaggeration to state that no other technology has offered so much promise as AI, nor so much risk.

Many of our observations below are the topics of active ongoing research. As such, we do not necessarily have the concrete answers for every question, but we wanted to share the direction that the research is heading. The national computing research community should be regarded as an ongoing resource with its unparalleled view of the leading edge of AI.

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;*
    - One of the greatest challenges is making sure that AI actors align with human values and norms to ensure that they behave ethically (an AI-related risk). In order to do this, AI actors need to incorporate complex ethical and commonsense reasoning capabilities to reliably and flexibly exhibit ethical behavior in a wide variety of interaction and decision-making situations.

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;

- Additional characteristics that should be considered in the Framework are evaluating the quality and trustworthiness of knowledge repositories from which AI systems are built, and the vulnerability of AI to intentional or inadvertent manipulation. It is important to improve knowledge repositories, resolve inconsistencies in the knowledge sources, and update the knowledge over time. In addition, while AI has the potential for transformative impacts across all sectors of society and the economy, there are concerns about the security and vulnerability of these systems. We know from experience that software is notoriously hard to debug, and AI will be even harder. AI systems also present a moving target as they are designed to adapt and learn through experience.

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;*
    - Additional important principles which should be considered in the Framework include beneficence, explainability, respect for human dignity and autonomy, and promoting equity and justice for all members of society.

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;*
    - The inability of researchers to access important proprietary data due to trade secrets is a concern for the computing research community. Normally academics play an important role in identifying applications of technology that present a danger to society. We can not do that, though, if we as computing researchers are not allowed access to critical data, or are prevented from speaking out by highly restrictive confidentiality agreements.

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;*
    - One critical AI risk management standard that NIST should consider is ensuring that an AI actor is a trusted human advisor and that it acts on behalf of the user, without the possibility of external manipulation. This includes blocking the acquisition of data that might impact the user negatively. Any risks should be accurately explained, and system uncertainty must be communicated to the user. This is a difficult situation to navigate, given the different purposes of AI systems,

but NIST needs to be mindful of the end user of a system and their needs and goals.

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.*
    - In order for AI design and development to be carried out in a way that reduces the negative impact on individuals, it is critical that AI algins with human values and norms. This is to ensure that they behave ethically and in our interests, hence the importance of diversity among developers of AI. Human society will need to enact guidelines, policies, and regulations that can address issues raised by the use of AI systems, such as ethical standards that regulate conduct. These guidelines must take into account the impact of the actions in the context of the particular use of a given AI system, including potential risks, benefits, harms, and costs, and will identify the responsibilities of decision makers and the rights of humans.

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.*
    - The Framework ought to be useful as guidance for curricular development, and should take input from academics with that particular goal in mind.
    - Much involving AI is active research, and will not know the (right) answers for some of these questions for years. That is why it is critical that there be Federal support for researchers who are blazing the trail and finding problems with AI before they actually hurt people.

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.*
    - The Framework should encourage open, collaborative, and interdisciplinary ecosystems that includes not only software developers, AI engineers, and computing researchers but also social scientists, ethicists, policy experts. etc. We encourage current best practices in software and AI development.