BlackBerry

September 7, 2021

National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899

Re: Artificial Intelligence Risk Management Framework

Dear NIST,

BlackBerry commends NIST's leadership and transparent, consensus-based process to develop a risk management framework for AI. We concur with NIST's objectives for the Framework to cover full-scale risks and provide a prioritized, flexible, and cost-effective approach that is scalable to organizations of all sizes, public or private, in any sector, and operating within or across domestic borders. We strongly support development of an AI RMF that provides a catalog of outcomes and approaches applicable for a variety of use cases, rather than a set of one-size-fits-all requirements, considering that:

- the catalog supports the prioritized, scalable and cost-effective objectives,
- the rigor and sophistication of AI risk management should be commensurate with the impacts of AI system outcomes to individuals, groups, society and organizations, and
- the relevance of the principles and characteristics for AI trustworthiness significantly varies depending on its intended use.

Below, we submit our response to the 12 questions included in the AI RMF RFI.

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;*

   The challenges include the lack of effective and flexible guidelines and tools for organizations to utilize to identify, access, prioritize and respond to risks in a cost-effective manner. In addition, there are no clear incentives and guidelines to communicate with stakeholders the AI risks that vary depending on multiple factors, including the intended use, maturity and scale of deployment.

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;*

   The listed AI characteristics are commonly recognized e.g. by policy makers, standards bodies, public and private sectors. We note that the relevance of each characteristic varies significantly depending on multiple factors, including the environment or sectors in which AI

BlackBerry Corporation
3001 Bishop Drive, Suite 400, San Ramon, California, 94583 USA. tel: +1 (925) 242-5660 fax: +1 (925) 242-5661

systems are deployed, the data they ingest, the functions or tasks they perform and the impacts of their outcome to individuals, society, environment and organizations. We understand that entities in critical infrastructure and regulated industries, e.g. healthcare and finance, have defined the trustworthiness characteristics essential to their use of AI and set their policies to manage them.

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;*

   We recognize the ten OECD AI principles are closely related to the NIST trustworthiness principles and characteristics. For example, the OECD value-based principles for AI actors share transparency, fairness and accountability with the NIST principles -- and privacy, explainability, robustness, security and safety with the NIST characteristics. We believe it would be helpful for NIST to explain how it selected three principles, classifying others as the characteristics, and how the principles and characteristics work together to achieve the goal of the Framework.

   It is imperative to establish clear definitions of the principles and characteristics of AI trustworthiness at the onset of AI RMF development. We note that ISO/IEC JTC 1/SC42 WG3 (trustworthiness) published a technical report ISO/IEC TR 24028 (overview of trustworthiness in artificial intelligence). The TR discusses most of the NIST AI trustworthiness attributes. We would ask that NIST consider the TR in developing its definition of AI trustworthiness attributes.

4. *The extent to which AI risks are incorporated into different organizations' overarching enterprise risk management – including, but not limited to, the management of risks related to cybersecurity, privacy, and safety;*

   The AI RMF is intended to be a tool that would complement and assist with broader aspects of enterprise risk management. Depending on the intended use of AI, outcome of the AI systems can impact cybersecurity of the organizations, privacy and safety of individuals. We agree that the AI RMF should be integrated with the organizations' overarching enterprise risk management. In this way, the organization can leverage and extend the existing policies, processes and organizational structure for cybersecurity, privacy and safety risk management and tailor the AI RMF optimally.

5. *Standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles to identify, assess, prioritize, mitigate, or communicate AI risk and whether any currently meet the minimum attributes described above;*

   Standards organizations including ISO, IEEE are currently developing guidelines or standards for AI risk management. For example, IEEE has the Applied Artificial Intelligence Systems (AIS) Risk and Impact Framework Initiative. ISO/IEC JTC 1/SC42 WG3 is

currently progressing ISO/IEC 23894 (AI – risk management) to the draft international standards stage and is targeting January 2023 for its completion. The standards intend to cover the minimum attributes described above.

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 European Commission published a memorandum on the proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) on April 21, 2021. The proposed regulation calls for AI standards, methodologies and tools for its support.

   The Regulation states that standardization should play a key role in providing technical solutions to AI providers to ensure compliance with this Regulation and that harmonized standards should be a means for AI providers to demonstrate conformity with the Regulation. For example, risk management measures shall take into account the generally acknowledged state of the art, including as reflected in relevant harmonized standards (Article 9). The proposed regulation also states that High-risk AI systems should perform consistently throughout their lifecycle and meet an appropriate level of accuracy, robustness and cybersecurity in accordance with the generally acknowledged state of the art. The level of accuracy and accuracy metrics should be communicated to users.

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;*

   Supervisory Guidance on Model Risk Management (SR11-7) was issued by the Federal Reserve and Office of the Comptroller of the Currency (OCC) in April 2011. Whilst the guidance primarily addresses banks' use of a model - a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates, its guidance on the model development, implementation, use and validations, and governance, policies and controls is applicable in many ways to risk management of AI systems that impact individuals, society and organizations. We think the guidance is useful for the AI RMF, considering the Framework aims to cover the full range of AI risks.

   We also recommend that NIST consider the OECD Framework for the classification of AI systems for the AI RMF in terms of risk identification, assessment and prioritization. The OECD framework defines four dimensions -- context, data and input, AI model, task and output -- and demonstrates they help organizations clearly identify and assess impacts of AI systems outcomes on human rights, democratic values and well-being by applying them to four significantly different use cases. Moreover, the outcome of the OECD's consultation shows that the implication of tasks performed by AI (e.g. forecasting) and relevance of the value-based AI principles or trustworthiness attributes vary significantly depending on the

application areas and industrial sectors. We propose that the OECD framework be considered as one of the key tools for the AI RMF to assist organizations to identify, assess and prioritize AI risks specific to their own use of AI. The OECD framework will support the objectives of the AI RMF, the framework covering full-scale risks, cost-effective and scalable to organizations of all sizes, public or private, in any sector, and operating within or across domestic borders, as described in the RFI.

For the standards currently under development, we recommend that NIST consider ISO/IEC 23894 (Artificial intelligence - Risk management), which provides AI specific guidelines on the basis of guidance given in ISO 31000:2018 (Risk management - Guidelines), and ISO/IEC 42001, which defines requirements and controls for the AI Management System. In general, we would request that NIST harmonize the AI RMF with international AI standards. There will be plenty of opportunities since both are currently under development.

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.*

IEEE stated that developing AI standards needs to be inclusive of diverse communities of experts and users, including economists, ethicists, legal professionals, philosophers, educators, policy-makers, regulators, and community representatives, in addition to technologists and scientists in its response to NIST RFI: Developing a Federal AI Standards Engagement Plan.

We agree that inclusiveness becomes more important in design, development, use and evaluation of AI systems as their impacts on the safety and well-being of individuals, society and the environment increase. Multi-disciplinary and domain experts' inputs in the early stages of AI system development, including the concept development, data planning (e.g. data collection and quality requirements) and design phases, contribute to reduce risks of unintentional, unanticipated, or harmful outcomes.

At the same time, the rigor and sophistication of the AI risk management should be commensurate with the impacts of AI system outcomes on individuals, groups, society and organizations. We would request that NIST develop the AI RMF to provide flexible guidelines for recommended inclusiveness based on the classification of AI systems e.g. by the OECD framework.

9. *The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes");*

We concur with NIST's approaches in every respect, as described in the AI RMF development and attributes section. At the onset of AI RMF development, it is imperative to agree on the common definitions as foundation of the development. NIST should provide definitions of AI trustworthiness principles and characteristics.

BlackBerry Corporation
3001 Bishop Drive, Suite 400, San Ramon, California, 94583 USA. tel: +1 (925) 242-5660 fax: +1 (925) 242-5661

We note that Clause 5.14 of ISO/IEC 22989 – Artificial intelligence concepts and terminology (currently in draft international standards stage) discusses the principles and characteristics of trustworthiness. NIST should consider harmonizing its definitions with the international standard, to the extent possible.

We strongly support the framework offering a catalog of outcomes and approaches to be used voluntarily. As demonstrated in the OECD framework for AI classification, the relevance of AI principles and implication of tasks AI performs vary depending on multiple dimensions, including the fields of AI applications, input data and tasks. Such catalogs of outcomes and approaches should enable organizations to tailor and implement the AI RMF in prioritized, flexible and cost-effective ways. Moreover, the AI RMF should be scalable to organizations of all sizes. We agree with NIST that a set of one-size-fits-all requirements will not achieve a flexible and cost-effective Framework. The catalogs of outcomes and approaches should be made available based on the AI system classification e.g. by the OECD framework. Finally, we request that NIST clarify the definition of the catalogs at the onset of AI RMF development.

10. *Effective ways to structure the Framework to achieve the desired goals, including, but not limited to, integrating AI risk management processes with organizational processes for developing products and services for better outcomes in terms of trustworthiness and management of AI risks. Respondents are asked to identify any current models which would be effective. These could include – but are not limited to – the NIST Cybersecurity Framework or Privacy Framework, which focus on outcomes, functions, categories and subcategories and also offer options for developing profiles reflecting current and desired approaches as well as tiers to describe degree of framework implementation;*

    We agree that the NIST Cybersecurity Framework and Privacy Framework provide effective foundation. Both frameworks are needed to form a basic structure for the AI RMF because AI impacts both organizations and individuals. Another efficient model is ISO management system standards (MSS). The High Level Structure allows multiple management system standards, e.g. ISO 27001 (information security management standard) and ISO 9001 (quality management standard) can be integrated into a single management system for organizations. Due to the global influence of NIST and ISO, we recommend that the AI RMF include mapping to ISO/IEC 23894 and 42001 in the future.

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 AI RMF should include guidelines for skills and competencies required for the roles and responsibilities that are required for organizations to implement successful AI risk management. The organizations tailor the guidelines to their needs, create and execute their own resource planning, which serve to advance the recruitment, hiring, development, and retention of a knowledgeable and skilled workforce.

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.*

    A governance framework should provide support and structure to risk management functions. For example, the framework can supply the organizations with high level guidance for risk management policies, processes to implement those policies and monitor the effectiveness of the processes. In this regard, we think that the AI RMF should include key aspects of the governance (e.g. risk tolerance and data governance) and risk management framework to support the governance requirements.

In sum, BlackBerry strongly supports NIST's leadership and approach to develop an AI RMF which covers full-scale risks and provides prioritized, flexible, scalable and cost-effective guidelines, as described in the RFI. We welcome the opportunity to offer our input; Mr. Takashi Suzuki tsuzuki@blackberry.com is available to respond to any questions about BlackBerry's response.

Respectfully submitted

Takashi Suzuki

Takashi Suzuki,
Senior Director, Standards

BlackBerry Corporation
3001 Bishop Drive, Suite 400, San Ramon, California, 94583 USA. tel: +1 (925) 242-5660 fax: +1 (925) 242-5661