SUBMITTED ELECTRONICALLY VIA REGULATIONS.GOV

August 19, 2021

Subject: Artificial Intelligence Risk Management Framework [Docket Number: 210726-0151]

U.S. National Institute of Standards and Technology
MS 20899, 100 Bureau Drive
Gaithersburg, MD 20899

To Whom It May Concern:

UL appreciates the opportunity to submit these comments to NIST on the Artificial Intelligence Risk Management Framework RFI. UL shares NIST's belief that even as AI presents great opportunities for advancement of the US economy and society at large, it also presents risks and challenges to that same society. AI must comport with ethical values, norms and legal expectations of specific societies or cultures, be designed, developed, used, and evaluated in a trustworthy and responsible manner, and minimize harms to individuals, groups, communities, and societies at large.

Since its inception in 1894, UL serves a mission of promoting safe living and working environments for people everywhere and fulfills a promise of facilitating the flow of goods across borders. Grounded in science and collaboration, UL's work empowers trust in pioneering technologies, from electricity to the internet. We help innovators create safer, more secure products and technologies to enable their safe adoption. Our nonprofit arm, Underwriters Laboratories, conducts rigorous independent research and analyzes safety data, convenes experts worldwide to address risks, shares knowledge through safety education and public outreach initiatives, and develops standards to guide safe commercialization of evolving technologies.

AI/machine learning (ML) enabled products present a unique challenge to product safety and safety certification. UL recognized this challenge immediately and has developed some practical elements of a risk management framework that may be relevant to NIST. These elements are now complete and their speedy introduction has helped facilitate a conversation with many stakeholders such as Automotive OEMs, startups and standards development organizations (SDOs), allowing us to continually to gather feedback and evolve our offerings.

In 2020, UL issued the first version of the ANSI/UL 4600, *Standard for Safety for Autonomous Products* (AP). The standard describes an approach that is a recognition of the challenge that AI/ML creates in bringing autonomous products to market. This standard follows a safety case assurance process as the usual prescriptive approach for product safety is not practical. The standard provides key elements that an autonomous products manufacturer must consider by gathering evidence to demonstrate the safety level of their product. Since the release of this first version, UL has provided formal training in UL 4600 as part of its UL Certified Autonomous Safety Professional program. (UL Certified Autonomy Safety | Professional Training | UL 4600 (kvausa.com)).

UL LLC
1850 M St. N.W., Suite 1000, Washington, DC 20036-5833 USA
T: 202.296.1435 / F: 202.872.1576 / W: UL.com

In addition, this year UL has established a new service, the AI Algorithm Reproducibility Process Verification Mark. This offering allows manufacturers to demonstrate their willingness to have their algorithm process evaluated by an independent third party. Reproducibility is a key element of trustworthiness, and this offering helps fill a gap in the standards for AI trustworthiness.

UL appreciates NIST's statement in the RFI regarding a preference for existing standards, which is consistent with the principles of Office of Management and Budget (OMB) Circular A-119, and that the AI RMF be developed in an open and transparent process. This development has brought together a variety of stakeholders including software developers, sensor and radar manufacturers, ride share companies, car manufacturers, state and federal agencies as well as consumer advocacy groups. In addition to ANSI/UL4600, Underwriters Laboratories has published ANSI/UL 5500, *Standard for Safety for Remote Software Updates.*

Please find below UL's responses to a subset of the questions posed in the Request for Information. As NIST moves forward with its efforts to develop an AI Risk Management Framework, UL is eager to share our valuable expertise with NIST. If you have any questions regarding this submission or would like to discuss UL's recommendations further, please do not hesitate to contact Thomas Daley, UL Global Government Affairs, at thomas.daley@ul.com. Thank you for your attention to these comments.

Sincerely yours,

David S. Wroth
Director, Data Science
Underwriters Laboratories Inc.

QUESTIONS FROM NIST

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;

UL RESPONSE: Understanding the malleability of AI systems: AI systems are heavily reliant on training data, sensor/input data feedback (reinforcement learning) and other external input that can and do alter system behavior and performance. The literature is full of examples where AI systems produce unexpected results due to the difficulty in fully understanding the patterns developed through neural networks and machine learning approaches. One example of this is the use of adversarial techniques, including small physical changes to "known" objects to produce unexpected and harmful behavior. One example is documented in "*Robust Physical-World Attacks on Deep Learning Models."*[1] Because systems may be more malleable than designers, users or managers recognize, it is difficult to fully understand

[1] Eykholt, et.al., Robust Physical-World Attacks on Deep Learning Models", arXiv:1707.08945v5 [cs.CR]

risk patterns. AI actors must consider "hard limits" to ensure AI systems don't operate in harmful ways when the system encroaches on the boundaries of the operational design domain.

Understanding the risk associated with a "system of systems": AI systems, complex in and of themselves, operate as part of other systems to be useful to society. Minimally, AI systems operate as part of a computer system, connected to a network system. Beyond that, the AI may be connected to sensor systems and control systems. These interconnections may be a source of cyber vulnerability or a source of unanticipated feedback or action. Modeling the complexity of the systems and their permutations is a challenge that many AI actors may underestimate.

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;

UL RESPONSE: Context is an important characteristic of trustworthiness. While trustworthiness is normally a positive attribute, there are situations where 'distrust' may be valuable. Distrust in the form of healthy skepticism of an AI system may encourage higher levels of safety, security and privacy. Defining the context of the system's use and outcome may be an important element of the user's ability to determine the level of trust put on the output of the AI. For example, in the medical field, the use of an AI in a diagnostic context requires high levels of trust by individuals and doctors to gain the benefits of the AI system. However, if an AI were used in defining a treatment program for a life-threatening condition, a level of 'distrust' would be useful to ensure corroboration by experienced doctors, similar to obtaining a "second opinion" which is common.

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;

UL RESPONSE: As covered in the introductory section of these comments, ANSI/UL 4600, *Standard for Safety for the Evaluation of Autonomous Products,* utilizes a "safety case" which is a structured explanation in the form of claims, supported by argument and evidence, that justifies that the item is acceptably safe for a defined operational design domain, and covers the item's lifecycle. The safety case:

- leads designers through the thought process required to consider the possible complications the system may encounter;
- requires evidence the system is sufficiently robust to mitigate foreseeable hazards;
- Classifies criteria as falling into Mandatory, Required, Highly Recommended, or Recommended; and
- provides prompts of "have you considered," examples and known pitfalls.

Furthermore, ANSI/UL 4600 requires that the safety case is reviewed by an independent, knowledgeable individual for completeness and evidence. Section 8.5.6 of UL 4600 specifically addresses the use of AI in autonomous products.

This approach could be expanded to address other dimensions of trustworthiness, for example by a companion “privacy case” or “fairness case.”

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;

UL RESPONSE: AS NIST may be aware, ISO / IEC Joint Technical Committee 1, Subcommittee 42 on Artificial Intelligence has developed standardization in the area of Artificial Intelligence, including the focus and proponent for JTC 1's standardization program on Artificial Intelligence, and provides guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications. Recent standards developed by this committee include:
ISO/IEC TR 24028:2020 Information technology -- Artificial Intelligence -- Overview of trustworthiness in artificial intelligence
ISO/IEC TR 24029-1:2020 Artificial Intelligence -- Assessment of the robustness of neural networks — Part 1: Overview
ISO/IEC TR 24029-1:2020 Information technology -- Artificial Intelligence -- Use cases

In addition to the work of ISO Joint Technical committee 1, standards in progress also include:
ISO/IEC 22989: Artificial Intelligence Concepts and Terminology
ISO/IEC 23053: Framework for Artificial Intelligence Systems Using Machine Learning
ISO/IEC 42001: Information technology -- Artificial Intelligence -- Management Systems
ISO/IEC 5259-1: Data quality for analytics and ML — Part 1: Overview, terminology, and examples
ISO/IEC 5259-2: Data quality for analytics and ML — Part 2: Data quality measures
ISO/IEC 5259-3: Data quality for analytics and ML — Part 3: Data quality management requirements and guidelines
ISO/IEC 5259-4: Data quality for analytics and ML — Part 4: Data quality process framework
ISO/IEC 24668: Information technology -- Artificial Intelligence -- Process management framework for Big data analytics
ISO/IEC Preliminary Work Item - Information technology – Artificial intelligence – Data life cycle framework
ISO/IEC TR 24027: Information technology -- Artificial Intelligence (AI) -- Bias in AI systems and AI aided decision making
ISO/IEC 24029-2: Artificial Intelligence (AI) -- Assessment of the robustness of neural networks -- Part 2: Formal methods methodology
ISO/IEC 23894 -- Information technology -- Artificial intelligence -- Risk management

ISO/IEC TR 24368: Information technology -- Artificial Intelligence (AI) -- Overview of Ethical and Societal Concerns
ISO/IEC TR 5469: Artificial Intelligence (AI) -- Functional Safety
ISO/IEC 25059 -- Software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality Model for AI-based systems
ISO/IEC TS 6254 -- Information technology -- Artificial intelligence -- Objectives and approaches for explainability of ML models and AI systems
ISO/IEC TS 5471 -- Artificial intelligence -- Quality evaluation guidelines for AI systems
ISO/IEC TR 24372: Information technology -- Artificial Intelligence (AI) -- Overview of computational approaches for AI systems
ISO/IEC TS 4213: Assessment of classification performance for machine learning models
ISO/IEC 5392: Information technology -- Artificial Intelligence (AI) -- Reference architecture of knowledge engineering
ISO/IEC AWI 38507 -- Information technology -- Governance of IT -- Governance implications of the use of artificial intelligence by organizations

NIST may consider joining the "Closing the Gaps in Responsible AI" effort by the Partnership on AI (www.partnershiponai.org). This initiative is a multiphase, multi-stakeholder project aimed at surfacing the collective wisdom of the community to identify salient challenges and evaluate potential solutions. These insights can in turn inform and empower the changemakers, activists, and policymakers working to develop and manifest responsible AI.