SUBMITTED ELECTRONICALLY VIA REGULATIONS.GOV

June 10, 2019

AI-Standards
National Institute of Standards and Technology
100 Bureau Drive, Stop 2000
Gaithersburg, MD 20899

Subject: Artificial Intelligence Standards [Docket No. 1903312229-9229-01]

About UL

UL is a premier global independent safety science company that has championed technological progress and societal well-being for 125 years by using research, standards, and conformity assessment to address important and emerging safety, security and sustainability challenges. Its nearly 14,000 professionals are guided by the UL mission to promote safe working and living environments. UL's functional safety and security activities and expertise grant us a broad view of the trends impacting connected technologies. This enables UL to anticipate and act upon technology implementation challenges, leading UL to establish IoT-related product standards and programs that fulfill its mission by bringing the science of safety to the forefront.

UL understands that the evolution of intelligent IoT (inclusive of applicable AI technologies) brings both new opportunities and risks to society. Through UL's safety science research and its various stakeholder services, UL ensures that connected products and systems work such that they are more secure, perform as intended, and are appropriately compatible with the connected environment to enable their safe use. UL has long been at the forefront of safety science. In doing so, UL has needed to anticipate, understand and then develop repeatable and reproducible means to assess new technologies. The objective – applicable to AI as well – has been to achieve safe outcomes that meet societal expectations for risk.

AI Introduces New Risks

AI represents the latest technical frontier to which safety science will be applied so that products and systems continue to provide safe outcomes. Technology is moving from discrete products, factory "programmed" to perform certain functions, to products that we, as users, could modify within an acceptable range determined by the product manufacturer. Connected (IoT) products represent a paradigm shift in that in-use products can be remotely modified (re-programmed) by authorized – and potentially unauthorized – entities applying a planned update.

AI extends the paradigm shift to include artificial (non-human) entities that take inputs from a number of sources (e.g. sensors) and apply machine logic (residing in the cloud, in edge devices or some combination of these) to modify products functioning alone or as a system. The machine logic is dynamic, continuously learning and, as complexity increases, risks a lack of transparency on how machine decisions are made. Well-executed and timely decision-making can, of course, lead to improved safety outcomes. One safety science challenge that AI therefore presents is how to differentiate, with confidence during assessment, that safe outcomes will be achieved in accordance with societal expectations.

Analyzing AI Introduces New Challenges

Software safety assessment validates that software under examination meets the intended purpose. It is typically accompanied by product/system verification testing (normal and abnormal operation conditions, including introduction of faults). While programmable electronics have for years been complex, it has still been practical to determine potentially unacceptable outcomes and to test for their potential occurrence. AI will likely require new approaches to validation of machine learning tools, the weighting and other parameters that yield probabilistic outcomes, measures to avoid bias and unsuitable inputs, and other decision-making boundary conditions. Similarly, when it may be impractical to identify non-deterministic outcomes, new approaches will be required to verify the technology yields safe outcomes meeting societal expectations.

AI Requires Emphasis on Safety

AI is already being incorporated into systems that have safety critical applications. The most noteworthy, of course, are autonomous systems that are being deployed in factories (autonomous material handling equipment), on our roads (autonomous vehicles) and in the air (unmanned aerial vehicles). Beyond these highly visible applications, safety critical AI are being deployed in a variety of industries, including healthcare, chemical and process control, maintenance and predictive analytics, and smart home applications. In each of these cases, the physical safety and security of people are already, and will increasingly be, based on the application of robust and trustworthy AI systems.

The NIST RFI discusses the need for robust, reliable and trustworthy AI systems. UL believes that that safety must be an explicitly stated criteria for the advancement of AI systems, and that specific standards should be developed for AI that will be incorporated into safety critical applications. While some might consider that a trustworthy system will inherently be safe, we anticipate, based on 125 years of safety engineering that, unless safety is an explicit part of the design, deployment and ongoing validation of AI systems, it may be underappreciated and will result in unintentional injury and loss of property and security.

As UL has assessed the AI landscape, we have seen that some frameworks for advancing artificial intelligence do not directly discuss the need for safety, but like NIST, incorporate some language around safety in the topic of "robust" AI[1], "prudent"[2] or "human-centric" AI[3]. The Partnership on Safer AI specifically recognizes the need to provide guidance on the safe application of AI through one of their "pillars" and working groups – Safety Critical AI. Similarly, the OECD recommends that AI specifically be safe, in addition to robust and secure[4].

We believe that it is critical to explicitly incorporate safety in AI standards development activity, as there are already many examples of the failure of AI that have, or could portend, safety issues. UL sponsored a literature review of AI system failures and compiled a list of more than 50 examples of real-world and simulation failures with potential safety, security or ethical implications.[5]

Furthermore, as AI is used to improve the efficiency and performance, these applications involving products combined into systems may operate such that they produce unintended consequences that have safety implications. Without frameworks and standards that focus on AI safety, machine learning algorithms and autonomous systems have the potential to change in ways that designers may not anticipate. As these systems push the boundaries of performance, we will see increasing risk of unsafe exploration of operating parameters, challenges to the scalability of supervision of AI systems and other manifestations of safety issues with potentially insufficient risk mitigation.[6]

In addition to the safety implications of unintended consequences, risks to safety are also introduced through intentional attacks on AI systems. The "attack surfaces" are vulnerable to a variety of vectors, some of which are unique to the AI space. The term "adversarial AI" has been coined to capture the range of issues associated with intentional efforts to fool, corrupt, bypass or otherwise negatively impact AI systems. While much of the work in adversarial AI has been done in academia, these approaches are now being demonstrated against deployed AI systems, such as against autonomous vehicle software.[7] Safer applications of AI systems will additionally need to consider adversarial AI attacks as well as the more traditional cybersecurity attacks on hardware and software that run the algorithms. AI frameworks should encourage standards that ensure appropriate attention is directed to secure implementation of AI in safety critical applications.

[1] Ethics Guidelines for Trustworthy AI, High Level Expert Group on Artificial Intelligence, European Commission, 2019. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477

[2] Montreal Declaration for a Responsible Development of Artificial Intelligence, 2018. http://montrealdeclaration-responsibleai.com

[3] A Proposed Model Artificial Intelligence Governance Framework, Personal Data Protection Commission, Singapore, January 2019. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf

[4] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449.

[5] *Current State of Knowledge on Failures of AI Enabled Products*, Yampolskiy, R., University of Louisville, January 2018

[6] Concrete Problems in AI Safety, Amodei, D. et. Al. 25 July 2016. https://arxiv.org/abs/1606.06565v1

[7] "Three Small Stickers in Intersection can Cause Tesla Autopilot to Swerve into Wrong Lane" IEEE Spectrum, 1 April 2019. https://spectrum.ieee.org/cars-that-think/transportation/self-driving/three-small-stickers-on-road-can-steer-tesla-autopilot-into-oncoming-lane

The characteristic of AI systems to rely on data to create, train and modify the algorithms requires additional considerations to ensure safety, ethical use and beneficial outcomes. These considerations include validation that training data sets are appropriate for the intended use of the system (representative of the real-world data the system will use), security of the data during compilation and processing to avoid injection of inappropriate or malicious data, and the ongoing validation of data streams, sensor input or other data sources during operation. AI frameworks should encourage standards to minimize the risk that invalid, inappropriate, incomplete or malicious data may cause safety implications in critical AI systems.

The Pace of Technology and its Impact on Standards

Product designs have evolved at an ever-increasing rate during the last several decades. The incorporation of firmware, software, and internet connections into products allows designers, manufacturers and operators to now update and modify products even after they have been put into operation. Standardization is keeping pace. UL has developed standards to address remote updates and to assist with the reduction of risk when the resulting product changes occur. Connected products have also introduced new security concerns, which are being assessed and mitigated through the application of UL standards, tests and certification methods. This experience logically will also be applied to new AI enabled products and systems.

However, it is foreseeable that AI-driven product change can be especially challenging. Machine learning algorithms, sensors and other data will enable the systems to “learn” through interaction with its environment and with the data it ingests. These algorithms will be processing volumes of data that challenge traditional validation methods, and they will also autonomously modify operating parameters of systems. This means that the component, product or system that is currently in use will potentially not behave in the same manner as the device that was originally designed, trained and put into operation. Standards employing new techniques and risk-mitigation strategies will be required to ensure that these ever-evolving, and highly complex systems remain with levels of acceptable risk to people and property.

UL has recognized that the pace of AI technology adoption and change requires an ever more agile approach to standards development without compromising safety expectations or unnecessarily delaying the benefits of the technology. To that end, experience is being gained through a number of ongoing standardization initiatives that will be quite beneficial in developing and refining our approach to AI standardization.

UL has developed numerous and diverse functional safety requirements for products and systems, software safety standards (UL 1998[8]), cybersecurity standards (UL 2900[9] series) and recently a remote software update standard (UL 5500[10]). UL’s breadth of knowledge and expertise, as represented by this range of standardization, has given us the perspective necessary to be able to consider the many factors that would come into play when developing AI-related standards. The

[8] UL 1998, *Software in Programmable Components*

[9] UL 2900, *Software Security for Network-Connectable Products*

[10] UL 5500, *Safety for Remote Software Updates*

experience has also led to a common foundation of terminology and concepts upon which each of these standards has been based.

Most recently, UL is in the process of developing a standard for the *Evaluation of Autonomous Systems* (UL 4600), the first draft of which will be released in July 2019. This standard will specifically address the ability of autonomous products to perform the intended function without human intervention using a goal-based approach citing principles and practices that must be addressed in creating a safety case. The standard provides a structured approach to highly complex and varied safety challenges introduced by AI. It is anticipated that UL 4600 primarily will be used in addition to functional safety product and system standards for autonomous products and can especially be applied to autonomous vehicle application.

Conclusion

UL applauds NIST's efforts to explore AI and emerging technologies and commends NIST for facilitating this initial public stakeholder engagement. The Federal government shares many of the same goals as the private sector, and we encourage relevant agencies to leverage the private sector's expertise, knowledge and resources when developing policies and administering programs. We encourage the government to utilize public-private partnerships in building policies by incorporating consensus-based standards, available accreditation schemes, and globally recognized practices to improve the safety of IoT products and systems. As NIST executes its responsibilities under the Executive Order, UL is interested in assisting NIST and other federal agency stakeholders, as appropriate, to further elevate the importance of safety in deployment of AI applications and lend our knowledge and expertise in research, standards development, and conformity assessment as needed.

UL appreciates the opportunity to provide these initial comments and looks forward to further engagement. In the meantime, if UL can be of further assistance or if you would like to discuss elements of this submission, please contact Abel Torres (abel.torres@ul.com).

Abel Torres
UL, Global Principal Policy Advisor