American Property Casualty
Insurance Association
INSURING AMERICA apci.org

October 15, 2020

National Institute of Technology
100 Bureau Drive
Gaithersburg, MD 20899

Via electronic delivery: explainable-AI@nist.gov

Re: Four Principles of Explainable Artificial Intelligence

To Whom it May Concern:

The American Property Casualty Insurance Association (APCIA) is pleased to provide feedback on the National Institute of Standards and Technology's (NIST) "Four Principles of Explainable Artificial Intelligence" (Principles Paper). APCIA is the preeminent national trade association representing property and casualty insurers doing business locally, nationally, and globally. Representing nearly 60 percent of the U.S. property casualty insurance market, APCIA promotes and protects the viability of private competition for the benefit of consumers and insurers. APCIA represents the broadest cross-section of home, auto, and business insurers of all sizes, structures, and regions of any national trade association.

Artificial intelligence (AI) has great potential to augment human judgment, improve decision making, and enhance the customer experience. Nevertheless, as with any emerging technology, there are issues that need to be considered and evaluated consistent with existing regulatory requirements. APCIA appreciates the high-level common-sense approach to explainable AI in the Principles Paper and offers additional observations for your consideration below.

Scope

As a practical matter, it may be beneficial for NIST to test the principles against real-life examples with technologists and AI practitioners to ensure they are realistic and attainable as outlined. In addition, there does not appear to be a definition of "AI systems." Is the paper solely focused on explainable AI for customer-facing or product-based efforts or are these types of explanations expected for every deployment of AI. In this respect, different audiences will necessitate different explanations.

Additionally, it should be recognized that explanations should not be necessary for every deployment of AI, such as low impact/low risk uses of AI. The degree of explainability with respect to a particular use of AI must be balanced leveraging the accuracy and efficiency that AI can bring. Moreover, even with high risk/high impact use of AI, there are some systems so complex that only very technical explanations can be provided, thus somewhat defeating the purpose of explanation.

Explanation Accuracy

Principle 3 highlights that it is hard to have both meaningfulness and accuracy in explainable AI. Accordingly, the Principles Paper would benefit from clarity on how NIST expects companies, such as insurance companies and other financial institutions, to incorporate these four principles into their systems. For instance, will there be a different level of expectation, based on the level of potential risk or impact of the AI? We encourage, NIST to consider providing some examples around this point.

Adversarial Attacks on Explainable AI

Section 5.4 points out that explanation accuracy can be exploited by adversaries. Specifically, the paper states that 100 percent explanation accuracy can be exploited by adversaries who manipulate a classifier's output on small perturbations of an input to hide the biases of a system and this in turn misleads users. Could NIST identify some recommended best practices or tools that can be utilized to minimize these risks?

Human Element

NIST's discussion of the fallibility of human explanations of their decisions is very important to consider in setting benchmarks for explainable AI. One of the challenges for AI is building consumer trust and eliminating fears around the lack of human intervention in decision making. The Principles Paper notes that humans unconsciously incorporate irrelevant information and personal biases into a variety of everyday decisions. If this is the starting point, there is an opportunity to leverage AI for societal good to correct, or at least help control, the shortcomings and biases of unconscious human decision making. To this end, the Principles Paper should further highlight these benefits of AI in relation to human decisions to help with the consumer trust issue while balancing reasonable expectations for AI in this area.

***

APCIA appreciates the opportunity to share our observations with NIST and looks forward to continued partnership and collaboration on AI issues.

Sincerely,

Angela Gleason

Angela Gleason
Senior Director, Cyber & Counsel