Risk management framework aims to improve trustworthiness of artificial intelligence

January 27 2023

Credit: N. Hanacek/NIST

The U.S. Department of Commerce's National Institute of Standards and

Technology (NIST) has released its Artificial Intelligence Risk Management Framework (AI RMF 1.0), a guidance document for voluntary use by organizations designing, developing, deploying or using AI systems to help manage the many risks of AI technologies.

The AI RMF follows a direction from Congress for NIST to develop the framework and was produced in close collaboration with the private and public sectors. It is intended to adapt to the AI landscape as technologies continue to develop, and to be used by organizations in varying degrees and capacities so that society can benefit from AI technologies while also being protected from its potential harms.

"This voluntary framework will help develop and deploy AI technologies in ways that enable the United States, other nations and organizations to enhance AI trustworthiness while managing risks based on our democratic values," said Deputy Commerce Secretary Don Graves. "It should accelerate AI innovation and growth while advancing—rather than restricting or damaging—civil rights, civil liberties and equity for all."

Compared with traditional software, AI poses a number of different risks. AI systems are trained on data that can change over time, sometimes significantly and unexpectedly, affecting the systems in ways that can be difficult to understand. These systems are also "socio-technical" in nature, meaning they are influenced by societal dynamics and human behavior. AI risks can emerge from the complex interplay of these technical and societal factors, affecting people's lives in situations ranging from their experiences with online chatbots to the results of job and loan applications.

The framework equips organizations to think about AI and risk differently. It promotes a change in institutional culture, encouraging organizations to approach AI with a new perspective—including how to

think about, communicate, measure and monitor AI risks and its potential positive and negative impacts.

The AI RMF provides a flexible, structured and measurable process that will enable organizations to address AI risks. Following this process for managing AI risks can maximize the benefits of AI technologies while reducing the likelihood of negative impacts to individuals, groups, communities, organizations and society.

The framework is part of NIST's larger effort to cultivate trust in AI technologies—necessary if the technology is to be accepted widely by society, according to Under Secretary for Standards and Technology and NIST Director Laurie E. Locascio.

"The AI Risk Management Framework can help companies and other organizations in any sector and any size to jump-start or enhance their AI risk management approaches," Locascio said. "It offers a new way to integrate responsible practices and actionable guidance to operationalize trustworthy and responsible AI. We expect the AI RMF to help drive development of best practices and standards."

The AI RMF is divided into two parts. The first part discusses how organizations can frame the risks related to AI and outlines the characteristics of trustworthy AI systems. The second part, the core of the framework, describes four specific functions—govern, map, measure and manage—to help organizations address the risks of AI systems in practice. These functions can be applied in context-specific use cases and at any stages of the AI life cycle.

Working closely with the private and public sectors, NIST has been developing the AI RMF for 18 months. The document reflects about 400 sets of formal comments NIST received from more than 240 different organizations on draft versions of the framework. NIST today released

statements from some of the organizations that have already committed to use or promote the framework.

The agency also today released a companion voluntary AI RMF Playbook, which suggests ways to navigate and use the framework.

NIST plans to work with the AI community to update the framework periodically and welcomes suggestions for additions and improvements to the playbook at any time. Comments received by the end of February 2023 will be included in an updated version of the playbook to be released in spring 2023.

In addition, NIST plans to launch a Trustworthy and Responsible AI Resource Center to help organizations put the AI RMF 1.0 into practice. The agency encourages organizations to develop and share profiles of how they would put it to use in their specific contexts. Submissions may be sent to AIFramework@nist.gov.

Provided by National Institute of Standards and Technology

Citation: Risk management framework aims to improve trustworthiness of artificial intelligence (2023, January 27) retrieved 26 April 2024 from https://techxplore.com/news/2023-01-framework-aims-trustworthiness-artificial-intelligence.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.