NIST AI INNOVATION LAB

Carrying Out Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI)

Why Did NIST Do the Work?

EO 14110 (October 30, 2023) charges multiple agencies – including NIST – with producing guidelines and taking other actions to advance the safe, secure, and trustworthy development and use of AI.

NIST has a proven track record of work in AI and in technology-based risk management guidance.

Next Steps

AI actors are encouraged to review, use the guidelines produced by NIST, and provide input for possible future revision of the documents.

Engage in NIST's currents and upcoming efforts to operationalize those guidelines, and their incorporations in international standards.

Contact

ai-inquiries@nist.gov

Learn More

Progress to Date

Under the Executive Order (EO) and based on extensive public engagement, the guidance **coordinated by the NIST AI Innovation Lab (NAIIL) includes**:

The AI RMF Generative AI Profile (NIST AI 600-1) can help organizations identify unique risks posed by generative AI and proposes risk management options that align with their goals and priorities. A companion to NIST's AI RMF, it describes 12 risks and actions organizations can take to manage them.

To reduce synthetic content risks, NIST developed a memo from the Secretary of Commerce to the White House identifying existing standards, tools, methods, and practices – and the potential development of further standards and techniques related to synthetic content risks. A detailed report is being finalized.

A Plan for Global Engagement on AI Standards (NIST AI 100-5) is designed to drive the worldwide development and implementation of AI-related consensus standards, cooperation and coordination, and information sharing.

Secure Software Development Practices for Generative AI and Dual-Use Foundation Models (NIST SP 800-218A) expands the Secure Software Development Framework (SP 800-218), recognizing that with generative AI systems, software can be compromised with malicious training data that adversely affects AI system performance.

Guidelines for Evaluating Differential Privacy Guarantees *(Initial Public Draft)* is intended to improve understanding about how to evaluate promises made (and not made) when deploying differential privacy, including for privacy-preserving machine learning. A final version is due out soon.

NIST's Materials Measurement Laboratory is working with key stakeholders to develop standards and best practices related to nucleic acid screening.

In addition to those documents, NIST's AI Safety Institute released for comment the initial public draft of guidelines on *Managing Misuse Risk for Dual-Use Foundation Models* (NIST AI 800-1). The draft outlines voluntary best practices for how foundation model developers can protect their systems from being misused to cause deliberate harm to individuals, public safety and national security.

10/9/2024