KPMG

Regulatory Alert

Regulatory Insights

August 2024

New AI Actions: White House Announcement; NIST, NTIA Guidance

KPMG Insights:

- ***Marking Time:*** *The White House announces completion of the required 270-day actions under the AI Executive Order on schedule.*
- ***Rush of Releases:*** *Commerce Department, through NIST and NTIA, continues to release principle-based frameworks/ guidance under the AI EO, with more reports and guidance to come.*
- ***Broad Coverage:*** *AI issuances span GenAI-related risk management and development frameworks, global plans for alignment, and best practices for managing and mitigating risks in foundation models.*
- ***Quick Turnaround:*** *NIST's rapid issuance of AI guidance demonstrates the swiftness of AI policy/regulation for both AI developers and deployers.*

Marking 270 days since the President signed the Executive Order (EO) on the Safe, Secure and Trustworthy Development of AI (see KPMG's Regulatory Alert, here), the White House announces new actions taken by federal agencies in response to the EO, including these releases from the Department of Commerce's National Institute of Standards and Technology (NIST) and National Telecommunications and Information Administration (NTIA):

1. Final Guidance: Artificial Intelligence Risk Management Framework: Generative artificial intelligence (GenAI) Profile (NIST AI 600-1)
2. Final Guidance: Secure Software Development Practices for GenAI and Dual-Use Foundation Models (NIST Special Publication (SP) 800-218A)
3. Final Guidance: Global Collaboration on AI Standards (NIST AI 100-5)
4. Draft Guidance: Managing Misuse Risk for Dual-Use Foundation Models (NIST AI 800-1)
5. Report: Dual Use Foundation Models with Widely Available Model Weights (NTIA Report)

Key features of these five publications are outlined on the following pages.

The White House announcement presents a summary table of the activities undertaken by federal agencies in response to the EO. Recent actions highlighted by the White House include the following (those with asterisks are covered in more detail below):

Agency	Action
Chief Data Officers Council	Developed initial guidelines for federal agencies to uphold the government's mandate for transparency in data while protecting against the potential misuse of such data to harm national security through the training of AI systems
Department of Commerce	Submitted a report to the White House outlining tools and techniques to reduce the risks from synthetic content
	Developed guidelines and resources to support the implementation of minimum risk management practices from OMB's policy on federal agencies' use of AI
	* Published a profile of the AI Risk Management Framework covering generative AI
	* Published a profile of the Secure Software Development Framework incorporating practices for generative AI and dual-use foundation models
	* Established new guidelines and best practices for AI Developers to prevent their systems from being misused to harm individuals or threaten safety and security, and for developers to increase transparency about their products
	* Published a plan for U.S. engagement abroad on AI standards
	* Prepared a report – drawing from extensive expert outreach and public comments – on the potential benefits, risks, and implications of dual-use foundation models for which the model weights are widely available, including related policy recommendations
	Published guidance on evaluating the eligibility of patent claims involving inventions related to AI technology
Department of Defense, Department of Homeland Security	Submitted reports to the President on results from pilots of new AI tools to identify and address vulnerabilities in critical government systems and software
Department of Education	Released a guide for developers of education technologies on designing safe, secure, and trustworthy AI tools for use in education
Department of Energy	Finalized and began to implement plan to develop and expand AI testbeds and safety evaluation tools
Department of Homeland Security	Developed a plan for multilateral engagements to encourage adoption of AI guidelines by critical infrastructure owners and operators
Department of Justice	Evaluated best practices developed for federal law enforcement agencies to hire professionals with technical skills and train professionals in the responsible use of AI
Department of State	Issued a Risk Management Profile for AI and Human Rights with guidance for governments, the private sector, and civil society worldwide on identifying and managing AI-related risks to human rights
Department of Veterans Affairs	Hosted two nationwide AI Tech sprint competitions
National Science Foundation	Engaged federal agencies on PETs adoption and launched the Privacy-preserving Data Sharing in Practice Program (PDaSP) to support critical work to apply, mature, and scale privacy-enhancing technologies (PETs) for specific use cases
	Launched an initiative to help fund researchers outside the federal government design and plan AI testbeds
National Security Council, Office of the White House Chief of Staff	Developed the first-ever National Security Memorandum on AI and presented it to the President for review

1. Final Guidance: AI Risk Management Framework: GenAI Profile. The AI RMF: GenAI Profile (NIST AI 600-1) is designed as a "companion resource" for users of the NIST AI Risk Management Framework 1.0 (AI RMF 1.0). It is intended to serve as both a use-case and cross-sectoral profile of the AI RMF 1.0 related to GenAI risk management, as well as a resource to assist companies in considering legal and regulatory requirements, and industry best practices for managing GenAI-specific risks.

In particular, the profile defines a group of twelve (12) risks that are unique to, or exacerbated by, the use of GenAI, and provides more than 200 actions to help govern, map, measure, and manage them. These risks include:

- Chemical, biological, radiological, or nuclear (CBRN) weapons information or capabilities
- Confabulation (e.g., "hallucinations" or "fabrications")
- Dangerous, violent, or hateful content
- Data privacy, particularly biometric, health, location, personally identifiable, or other sensitive data
- Environmental impacts due to resource utilization in training GenAI models
- Harmful bias and homogenization
- Human-AI configurations (arrangement or interaction of humans and AI systems which may result in such things as "algorithmic aversion", automation bias, or misalignment of goals)
- Information integrity
- Information security
- Intellectual property
- Obscene, degrading, and/or abusive content
- Value chain and component integration (non-transparent/ untraceable integration of upstream third-party components (e.g., data acquisition and cleaning, supplier vetting across the AI lifecycle)

2. Final Guidance: Secure Software Development Practices for GenAI and Dual-Use Foundation Models. The Secure Software Development Practices for GenAI and Dual-Use Foundation Models (NIST SP 800-218A) is designed as a "Community Profile" companion resource to supplement NIST's existing Secure Software Development Framework (SSDF) (SP 800-218) and is intended to be useful to producers of AI models, producers of AI systems that use those models, and acquirers of those AI systems.

While the SSDF focuses on assisting companies to secure software's lines of code, the final guidance expands on that focus to help address concerns around malicious training data adversely affecting GenAI systems. The guidance adds practices, tasks, recommendations, considerations, notes, and other information specific to GenAI and dual-use foundation model development throughout the software development lifecycle, including potential risk factors (e.g., signs of data poisoning, bias, homogeneity, or) and strategies to address them.

3. Final Guidance: Global Collaboration on AI Standards. NIST's Plan for Global Engagement on AI Standards (NIST AI 100-5) calls for a coordinated effort to work with key international allies and partners and standards developing organizations to drive global development and implementation of AI-related standards, cooperation and coordination, and information sharing. The plan outlines recommendations in the areas of:

- Standardization: Priority topics (e.g., terminology/taxonomy, risk measurements/mitigations, shared practices for testing, evaluation, verification, and validation (TEVV) of AI systems, mechanisms for enhancing awareness of the origin of digital content (authentic or synthetic), etc.), risk-based management frameworks, and other topics that may require more scientific research and development to understanding about critical components of a potential standard (e.g., energy consumption of AI models, incident response and recovery plans).
- Collaboration: Prioritize engagement with international standards developers, particularly on research and related technical activities; facilitate diverse multi-stakeholder engagement, including private sector leadership/efforts both domestically and more broadly; promote international exchange and alignment on standards and frameworks.

As called for by the EO, NIST will report to the President on priority U.S. government actions undertaken pursuant to the plan within 180 days from publication.

4. Draft Guidance: Managing Misuse Risk for Dual-Use Foundation Models. NIST's U.S. AI Safety Institute releases draft guidance for Managing Misuse Risks for Dual-Use Foundation Models across the AI lifecycle. The guidance outlines seven (7) objectives for organizations (and initial developers in particular) to map, measure, manage and govern the risk that foundation models will be misused to deliberately harm security or public health or safety. These objectives, along with associated practices to help achieve them, include:

- Anticipating potential misuse risks.
- Establishing plans for managing misuse risk.
- Managing the risks of model theft.
- Measuring the risk of misuse.
- Ensuring that misuse risk is managed before deploying foundation models.
- Collecting and responding to information about misuse after deployment.
- Providing appropriate transparency about misuse risk.

Comment Period. The AI Safety Institute is soliciting public comments on the draft guidance (NIST AI 800-1), with a deadline to submit by September 9, 2024.

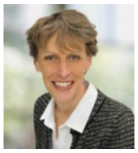
5. Report: Dual Use Foundation Models with Widely Available Model Weights. The Department of Commerce's National Telecommunications and Information Administration (NTIA) issues a report on Dual Use Foundation Models with Widely Available Model Weights. The report outlines several areas of risks and benefits to these "open-weight models", including public safety, geopolitical considerations, societal risks and well-being, and competition, innovation, and research.

The report details various policy approaches, such as restricting open-weight model availability, continuous monitoring and evaluation, and acceptance/promotion of openness, and recommends the following government actions:

- *Collect evidence* about the capabilities, risks, and benefits of the present and future ecosystem of dual-use foundation models with widely available model weights.
- *Evaluate evidence* by comparing indicators against specified thresholds, to determine when risks are significant enough to change the federal government's approach to open-weight foundation model governance (and when appropriate).
- *Act on evaluations* by adopting policy and regulatory measures targeted appropriately across the AI value chain.

For more information, please contact Amy Matsuo or Bryan McGowan.

Contact the author:



Amy Matsuo
Principal and National Leader
Regulatory Insights
amatsuo@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us: in **kpmg.com**