Submission of the Software & Information Industry Association

Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence, NIST-2023-0309 (889 FR 88368)

February 2, 2024

The Software & Information Industry Association (SIIA) welcomes the opportunity to provide input to National Institute of Standards and Technology (NIST) in carrying out several of its responsibilities under the Executive Order on *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, issued on October 30, 2023 (EO 14110).

SIIA is the principal trade association for companies in the business of information. Our members include nearly 400 companies reflecting the broad and diverse landscape of digital content providers and users in academic publishing, education technology, and financial information, along with creators of software and platforms used by millions worldwide, and companies specializing in data analytics and information services.

1. Developing Guidelines, Standards, and Best Practices for AI Safety and Security

1.a. Guidelines and best practices to promote consensus industry standards

1.a.1. AI RMF companion resource for generative AI

The NIST AI Risk Management Framework (AI RMF) has become a useful tool for industry and government alike in guiding the development and use of AI technologies in a trustworthy manner. SIIA has appreciated the collaborative manner in which NIST undertook development of the AI RFM. We believe this approach has only strengthened the reception and uptake of the AI RMF as the key cross-sectoral benchmark for assessing the safety, security, and trustworthiness of AI technologies.

Developing an AI RMF companion resource for generative AI (GenAI) will be a welcome addition to AI governance and the development of consensus industry standards. As a preliminary matter, we recommend that NIST identify appropriate standards and frameworks that already exist and focus the GenAI companion resource on memorializing those and filling in critical gaps - those that align with the potential risks associated uniquely with GenAI tools. In terms of existing standards, we recommend relying on the recently published ISO/IEC 42001 standard where appropriate.

In addition, we recommend NIST engage with those in industry who have begun to wrestle with how to mitigate the risks associated with GenAI. Several SIIA member companies are on the leading edge

of GenAI development and responsible practices.[1] And there has been significant work undertaken by multi-stakeholder organizations that will help to raise the bar on GenAI safety and security and vet tools across a range of models. We note, for example, the work being undertaken by ISO/IEC JTC 1/SC42, as well as that of the Partnership on AI, the OECD, and MLCommons.

We do recommend that as NIST proceeds, attention is given to distinguishing risk management guidance for different actors in the AI value chain. The Partnership on AI's Model Deployment Guidance, which remains open for public comment, can be useful for building a GenAI compendium focused on GenAI developers. However, as recognized in the current AI RMF, the risk mitigation steps appropriate for developers will differ from those appropriate for deployers. And in the GenAI context, attention to end users (including, for appropriate GenAI models, the general public) is critical. There are already countless GenAI tools that are being used in non-public ecosystems. The potential risks associated with these bespoke, closed systems differ materially from tools that are widely available to the public.

While the current AI RMF reflects core guidance for actors across the value chain, we believe NIST can provide significant value by vetting existing guidance and best practices to calibrate the AI RMF companion to the unique risks of GenAI and risks that GenAI tools may amplify, including generation of synthetic content and hallucination.[2] NIST's focus on context in the original AI RMF – "AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior" – is perhaps more fundamental when GenAI tools are considered. Language uses, community context, and cultural variations, for example, are important to the trustworthiness and efficacy of certain GenAI models. In addition, with regard to risk management as applied to end users, including the general public, one of the best risk mitigation strategies may involve digital literacy. We hope the forthcoming National AI Literacy Day will provide an opportunity to kick-start this type of effort.

Impact assessments have been discussed as a tool to offset the development of potentially harmful technologies. We believe that impact assessments prior to deployment are best practice and that continued monitoring for generative AI models should be encouraged. Additionally, GenAI tools that are intended to have an impact on actions affecting individual's legal rights or access to essential services should be subject to heightened assessment requirements.

The RFI also seeks input on content authentication and provenance tracking. These are important tools that we anticipate will become increasingly integral to GenAI systems as the technology matures. We have consolidated our feedback on this topic below.

---

[1] We recommend the following resources, among others, that address responsible development and use of GenAI models: Google, 2023 AI Principles Progress Update; Google, Evaluating social and ethical risks from generative AI (Oct. 2023); Meta, Introducing Purple Llama for Safe and Responsible AI Development (Dec. 2023); Meta, Llama 2 Responsible Use Guide; Amazon Web Services, Tools and Resources to Build AI Responsibly.

[2] See, e.g., State of California, Benefits and Risks of Generative Artificial Intelligence Report (Nov. 2023).

1.a.2. Guidance and benchmarks for evaluating and auditing AI capabilities

While there has been substantial work undertaken by academics, researchers, and industry, TEVV methods for GenAI remain in their infancy. We support efforts by NIST, NSF, and other agencies to foster guidance that can be used by developers and deployers to advance auditing, evaluations, and testing of GenAI data and models. We recommend seeking guidance from some of the organizations that have already begun to grapple with this challenge, including ISO, the OECD, the Partnership on AI, and MLCommons.

1.b. Red Teaming

Red teaming is an important component of robust TEVV processes across the AI lifecycle. We are encouraged by the uptake of red teaming across industry, both internally within organizations and in collaboration with the public, as a component for assessing the performance of GenAI systems. We believe there is value in NIST developing guidance on red teaming in the context of GenAI.

Red teaming should be considered part of a more holistic, comprehensive approach to GenAI risk management; while an important tool, it cannot substitute for other effective risk management processes. Adversarial attack simulation, for example, is another important tool that could be considered distinct from "red teaming" depending on how this is ultimately defined.

We recommend that NIST develop guidance focused on the performance of GenAI models. This approach would be consistent with EO 14110's definition of "red teaming" as a method "to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with the misuse of the system." We further recommend avoiding a "checklist" approach to red teaming, which could be constraining and could focus too much on isolated inputs or outputs that are too attenuated from systemic risks within the models. Highlighting best practices and featuring case studies of these would be helpful to raise the bar within industry.

2. Reducing the Risk of Synthetic Content

SIIA supports NIST's efforts to reduce the risk of synthetic content in both closed and open models. This is an area that has rightfully received significant attention from policymakers. We note, however, that the state of synthetic content identification and labeling remains in an early stage. We recommend that NIST consider a wide array of policy interventions that include both technical approaches and user education.

With respect to technical approaches, we note that there are promising efforts underway in industry, academia, and in government.[3] Among these, we encourage NIST to engage with the Semantic Forensics (SemaFor) team at DARPA. SemaFor has devoted years to developing algorithms and tools to detect synthetic media, some of which could be scalable.

[3] See, e.g., Partnership for AI, Building a Glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure (Dec. 19, 2023).

While there exist a range of tools to authenticate content or demonstrate provenance, there are limitations in these tools. Watermarking, for example, can limit the usability of certain content and can be removed or replicated by sophisticated actors. Furthermore, the efficacy of watermarks depends on wide adoption, which can limit their value in providing users with reliable information about content. Watermarking techniques alone are not enough to ensure the security and trust in the content we consume but built in as one tool with other provenance techniques, can be helpful in minimizing the spread of synthetic content.

Moreover, because one goal of GenAI tools is to generate synthetic content, the use of watermarks or labels cannot currently distinguish between content that is synthetic and benign and content that is synthetic and misleading. That effort requires continued attention by trust and safety teams across industry as well as a more informed public.

As research and technology develop, it will be important to couple technical solutions with digital literacy. An educated public is an essential component of a holistic approach to addressing the proliferation of synthetic media. We would encourage the U.S. government to create a program to support digital literacy, focused on GenAI, both for youth and adults. This would help to enhance public trust in the consumption of online media. NIST, for its part, could assist this effort in developing risk profiles for different types of synthetic media and different actors across the AI value chain.

We also note that an evaluation of the risk of synthetic content should also consider the applicability of federal and state laws to provide remedies for individuals who have been aggrieved by synthetic media.[4] Here is an existing, technology-neutral set of legal protections for individuals, and we understand Congress is actively considering legislation to address some potential harms left unaddressed by current law.

3. Advance Responsible Global Technical Standards for AI Development

SIIA supports efforts underway to align rules and policy internationally. We are pleased by the efforts of the G7 and the first-ever International Code of Conduct, as well as ongoing standards efforts by ISO/IEC JTC1/SC42 and other bodies. Global technical standards must be developed to create baselines for safety, security, and trustworthiness, while also allowing innovation to continue.

We believe NIST has had a critical role in the development of global alignment on AI standards. Its efforts to align taxonomy and definitions with EU counterparts, its participation in international technical standards bodies, and its work with the OECD, among others, are helping to foster AI that is safe, secure, and trustworthy.

We encourage further attention to involving smaller enterprises in international standardization efforts. Raising the bar on AI trustworthiness will benefit from having both large and small AI

[4] See Testimony of Chris Mohr, Software & Information Industry Association, to the House Committee on the Judiciary, Subcommittee on Courts, Intellectual Property and the Internet, "Artificial Intelligence and Intellectual Property, Part II – Identity in the Age of AI" (Feb. 2. 2024), at 4-9.

companies involved. Likewise, the conditions for safe, secure, and trustworthy innovation will be strengthened if newer or smaller entrants can build responsible AI systems in accordance with international standards. The NAIRR pilot program can contribute to this effort, and if successful, expanding the program to include an international research resource could contribute to U.S. efforts to shape the rules of the road for AI's future.

* * *

SIIA thanks NIST for the opportunity to provide input on these issues. We look forward to continuing to work with NIST in the next phase of its work on safe, secure, and trustworthy AI.

Sincerely,

Paul Lekas
Senior Vice President, Global Public Policy & Government Affairs
Software & Information Industry Association