**BSA Comments on the National Institute of Standards and Technology's Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence**
**February 1, 2024**

BSA appreciates the opportunity to provide comments on the National Institute of Standards and Technology's (NIST) Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence (RFI).

BSA is the leading advocate for the global software industry.[1] BSA members are at the forefront of developing cutting-edge services — including AI — and their products are used by businesses across every sector of the economy.[2] For example, BSA members provide tools including cloud storage and data processing services, customer relationship management software, human resource management programs, identity management services, and collaboration software. BSA members are on the leading edge of providing AI-enabled products and services. As a result, they have unique insights into the technology's tremendous potential to spur digital transformation and the policies that can best support the responsible use of AI.

BSA's views are informed by our experience working with member companies to develop the BSA Framework to Build Trust in AI,[3] a risk management framework we published more than two years ago to help companies mitigate the potential for unintended bias in AI systems. Built on a vast body of research and informed by the experience of leading AI developers, the BSA Framework outlines a lifecycle-based approach for performing impact assessments and highlights corresponding best practices.[4] Our experience on these issues informs our recommendations below.

---

[1] BSA's members include: Adobe, Alteryx, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, Hubspot, IBM, Informatica, Kyndryl, MathWorks, Microsoft, Okta, Oracle, PagerDuty, Palo Alto Networks, Prokon, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, TriNet, Twilio, Workday, Zendesk, and Zoom Video Communications, Inc.

[2] *See* BSA | The Software Alliance, Artificial Intelligence in Every Sector, *available at* https://www.bsa.org/files/policy-filings/06132022bsaaieverysector.pdf.

[3] *See* BSA | The Software Alliance, Confronting Bias: BSA's Framework to Build Trust in AI, *available at* https://www.bsa.org/reports/confronting-bias-bsas-framework-to-build-trust-in-ai.

[4] BSA has testified before the United States Congress and the European Parliament on the Framework and its approach to mitigating AI-related risks. *See, e.g.,* Testimony of Victoria Espinel, Public Hearing on AI & Bias, Special Committee on Artificial Intelligence in a Digital Age, European Parliament, Nov. 30, 2021, *available at* https://www.europarl.europa.eu/cmsdata/244265/AIDA_Verbatim_30_November_2021_EN.pdf; Testimony of Victoria Espinel, The Need for Transparency in Artificial Intelligence, Before the Senate Committee on Commerce, Science, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, *available at* https://www.bsa.org/files/policy-filings/09122023aitestimonyoral.pdf.

200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org

The RFI identifies a broad range of issues that are important to the safe and trustworthy development of AI, including risk management, content authenticity, red teaming, and the development of international standards. We commend NIST for its consultative approach to engaging with stakeholders on these issues and encourage the agency to continue creating engagement opportunities as it implements aspects of the recent executive order. Our comments below recommend:

- Supporting AI risk management by ensuring developers and deployers of high-risk AI systems perform impact assessments based on their role;
- Distinguishing between the different actors in the AI ecosystem;
- Leveraging existing industry standards on content authenticity and provenance:
- Promoting internal red teaming for high-risk AI systems; and
- Supporting the development of consensus-based international standards on AI by leveraging bilateral and multilateral fora.

## I. AI Governance

AI governance plays a critical role in establishing corporate safeguards and accountability mechanisms for the development and deployment of trustworthy AI. Several approaches, such as performing impact assessments for high-risk AI systems and distinguishing among the different actors in the AI value chain, can strengthen risk management efforts.

### A. The NIST AI Risk Management Framework identifies important practices to identify and address AI risks.

NIST has been particularly focused on risk management, and BSA has strongly supported its development of the AI Risk Management Framework (RMF). The RMF has contributed significantly to the establishment of practices that organizations can adopt to help identify and mitigate AI risks. The RMF is a flexible framework that highlights key areas that organizations should address, including identifying metrics for risk measurement and evaluation, delineating roles and responsibilities, assessing the AI system's trustworthiness characteristics, and establishing feedback processes. A key benefit of the RMF is that it creates a common language for organizations handling AI risks. For example, if a set of organizations implements risk management practices based on the RMF, those entities can more readily manage risks across their organizations because they share a common approach to risk management. NIST's accompanying RMF playbook helps operationalize these concepts. Further, industry profiles help illustrate how the RMF can be applied across sectors.

### B. Impact assessments are an important accountability tool for high-risk AI systems and should be conducted based on a company's role in the AI ecosystem.

The RMF also highlights the utility of impact assessments. Impact assessments are important accountability tools that help developers and deployers identify and mitigate risks associated with high-risk AI systems. The principal value of an impact assessment is that it allows an organization to rigorously examine its practices, which drives change in internal processes. These changes help organizations adapt to new and emergent risks and implement changes across their products and services. The fact that assessments are being performed for high-risk AI systems also promotes trust for external stakeholders because they will know that an organization is conducting a thorough examination of AI



200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org

systems, and that the assessments are available to regulators upon request in the event of an investigation.

Impact assessments should focus on high-risk AI systems, to ensure that organizations devote resources to addressing systems that pose the greatest potential risks. An AI system may be high-risk if it makes consequential decisions that determine an individual's eligibility for and result in the provision or denial of housing, employment, credit, education, access to physical places of public accommodation, healthcare, or insurance. AI systems are used in a wide range of scenarios that do not present such risks, from detecting and lowering background noise on a video call to optimizing manufacturing production. For low-risk systems — like an AI system used to predict the types of fonts used in a document — an impact assessment is not necessary. But for high-risk systems, developers and deployers should perform impact assessments to assess and mitigate risks. Importantly, there is no "one-size-fits-all" approach to evaluating and mitigating risks of AI; impact assessments should be tailored to address the nature of the system at issue and the type of harms it may pose. Data protection impact assessments are common in the field of privacy and can be leveraged to address overlapping issues in an AI context.

Organizations must also conduct impact assessments that reflect the risks of their specific AI system and their role in developing or deploying that system. Both developers and deployers should conduct impact assessments of high-risk systems — but those assessments must reflect their different roles. Because a developer is the entity that designs, codes, or produces an AI system, and a deployer is the entity that uses an AI system, these two organizations will have different roles in identifying and mitigating potential risks. Moreover, the two types of organizations will have access to different types of information — and will be positioned to take different steps to mitigate potential risks. For example, developers that design an AI system are well-positioned to have access to information about the type of data used to train the AI system, the system's known limitations, and its intended use cases. In contrast, a deployer using an AI system is well-positioned to have access to information regarding the specific ways in which it uses that system that impacts consumers. Any policies focused on supporting AI accountability should reflect these different roles and assign obligations accordingly.

### C. AI policies should reflect the different roles of different entities along the AI value chain.

As NIST addresses AI risk management issues, it should recognize that there are often a broad set of actors involved in developing and deploying an AI system. These actors must work together for the system to function properly — and to appropriately manage the risks associated with that AI system.

As described above, these different roles include both the developers and deployers of an AI system. Communication among these different actors is important to ensure the successful operation of accountability frameworks across the lifecycle of an AI system. Developers that design a high-risk AI system should provide deployers using that AI system with the information reasonably necessary for the deployer to conduct an impact assessment. This may include the AI system's capabilities, known limitations, and guidelines for intended use. By providing this information, a deployer can then assess the use of an AI system in light of the developer's intended use for the system and its known limitations. At the same time, because developers will not have insight into the actual use of the AI system and do not have a relationship with the consumer or end user, a deployer



200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org

should be responsible for monitoring issues that arise in downstream implementation, including facilitating feedback to identify such issues.

Organizations may also take on other roles, such as integrating an existing AI model into the organization's products and services. Any obligations placed on these organizations should similarly reflect their role in integrating the AI system into the organization's products and services.

Creating role-based obligations is not unique to AI; role-based responsibilities are considered best practice in privacy and security legislation worldwide.

## II. Watermarking and Content Authenticity

The RFI seeks input on a range of issues about labeling and detecting AI-generated content. It focuses on both techniques like watermarking — which can label AI-generated content, so that users know that an image or video was created using AI — and on standards for authenticating content and tracking its provenance — which can help users know when an image is real, and when and how it was edited.

Watermarking is an important technique for labeling AI-generated audio and visual content, because it can ensure users know that the content was created by AI. However, it is not the only tool that can enhance consumer understanding of when audio or visual content has been created by AI. Other technical measures that enable organizations to indicate whether content was created by their AI system may also be useful.

Tools that help to authenticate audio and visual content and provide information about its source and history also help promote AI accountability. BSA supports the Content Authenticity Initiative's (CAI) efforts to promote the open Coalition for Content Provenance and Authenticity (C2PA) standard for content authenticity and provenance. C2PA's goal is to build technical standards that help users understand who created an image, and how, when, and where it was edited. In effect, it will create a stamp of authenticity that can help consumers decide what audio and visual content is trustworthy and promote transparency around the use of AI. In conjunction with tools like watermarking, the use of such standards provides secure, indelible provenance.

Importantly, the government should look to leverage existing standards, like those developed by C2PA, rather than recreating those standards. This ensures that standards are developed through voluntary, consensus-based processes in consultation with a wide variety of stakeholders. This approach can continue to promote trust and confidence in new technologies.

## III. Red Teaming

Red teaming also plays an important role in enhancing the security and trustworthiness of AI systems — but the resources and costs involved in red teaming make it most appropriate for AI systems that pose significant risks. Robust testing and evaluation of these high-risk AI systems for safety, security, accuracy, and fairness is critical. Red teaming should be viewed as one of many tools that can be useful to ensure AI systems that pose significant risks are rigorously tested to surface potential problems. Using red teaming in this targeted way can help enhance responsible AI development. Importantly, red teaming can implicate confidential or proprietary information and, as a result, policies addressing red teaming should acknowledge that it can be performed internally within an



200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org

organization. We also highlight below three approaches that are important to incorporate when implementing red teaming.

First, policies on AI red-teaming should focus not only on efforts to identify security vulnerabilities, but also on efforts to interrogate a system for other failures that could arise in the AI context, including the generation of harmful content. Importantly, red teaming should also include probing the AI system to uncover biases. Together, this expanded focus can provide a comprehensive view of the threat landscape.

Second, red teaming is not necessarily a "one and done" exercise. Given that AI systems are constantly changing, and that technologies like generative AI are designed to produce different outputs, red teaming that probes different aspects of the system may be appropriate. Red-teams assessing generative AI systems and models may often need to continue exploiting the system or exfiltrating data until they have fully assessed the root cause of the cyber vulnerability or adversarial attack vector. In these cases, this type of cyber testing may lead to risks like data poisoning or changes in model behavior. Red-teams should ensure this testing is done in an isolated environment to avoid an impact on operational AI models.

Third, mitigations should take a defense-in-depth approach. Like security, where a range of technical mitigations are required to address vulnerabilities, mitigating AI risks may require implementation of a variety of techniques. For example, mitigations could range from including classifiers that designate harmful content to using a metaprompt to shape behavior. Effective red teaming enables the identification of system failures and helps facilitate a fulsome approach to developing mitigation strategies.

## IV. Development of International Standards

The US government has a long-standing history of supporting industry-led, consensus-driven efforts to develop international standards on a variety of technology issues.[5]

It is important to continue this approach with respect to AI. NIST should work closely with allies and partners to support the development of international standards that multiple governments can look to when crafting AI policies. This approach is important, because creating country-specific standards can impede global harmonization. NIST should leverage ongoing work in bilateral and multilateral fora, such as the EU-U.S. Trade and Technology Council and the G7, to support the work of international standards development organizations.

Notably, the International Organization for Standardization (ISO) and the International Electrotechnical Commission's (IEC) joint technical committee 42 has several work streams to develop standards on AI. In December, it published ISO/IEC 42001, the first global AI management system standard, which, among other things, addresses ethical issues and transparency and provides guidance for organizations on how to manage AI risks. NIST should strengthen its collaboration with technical experts in other countries to contribute to ISO's ongoing work, including its efforts to develop standards for testing AI systems.

* * *

[5] *See* OMB Circular A-119: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities, *available at* https://www.nist.gov/system/files/revised_circular_a-119_as_of_01-22-2016.pdf.



200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org

We appreciate the opportunity to provide comments on NIST's RFI and remain available as a resource as you continue to implement the AI executive order.



200 Massachusetts Avenue, NW
Suite 310
Washington, DC 20001

P 202-872-5500
W bsa.org