Consumer Technology Association™

1919 S. Eads St.
Arlington, VA 22202
703-907-7600
CTA.tech

February 2, 2024

Consumer Technology Association
Comments on
National Institute of Standards and Technology's Request for Information Related to NIST's Duties Under the Biden Administration Executive Order on Artificial Intelligence

The Consumer Technology Association® ("CTA")®[1] respectfully submits these comments in response to National Institute of Standards and Technology's ("NIST") Request for Information ("RFI") on several of NIST's responsibilities under Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI), issued on October 30, 2023 ("AI Executive Order").[2] These comments focus on the RFI's request for input on issues raised under Sections 4.1, 4.5 and 11 of the AI Executive Order.[3]

CTA supports NIST's effort to build a robust record and solicit input from industry, academia, and civil society on the many significant issues raised by the Administration's directives to NIST in the AI Executive Order. The scope of these issues, ranging from developing risk management principles for generative AI to establishing a global plan for engagement on consensus AI best practices, is significant. CTA and its members look forward to engaging with NIST and its federal agency partners to support this effort.

In response to NIST's request for information, CTA urges NIST to proceed with caution in implementing its responsibilities under the AI Executive Order when considering guidelines and best practices that may be necessary for the ecosystem of entities developing and deploying generative AI systems. To that end, CTA recommends that NIST take a holistic approach to AI safety, security, and privacy. AI as a category of technologies is not new, but generative AI systems and technologies such as large language models (LLMs) that underlie generative AI systems are emerging technologies that are evolving rapidly, as are the varied use cases and applications to which these technologies can be applied. Thus, NIST should consider framing

[1] CTA® is the tech sector. Our members are the world's leading innovators—from startups to global brands—helping support millions of jobs. CTA owns and produces CES®—the largest, most influential tech event on the planet.

[2] The White House, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (Oct. 30, 2023) https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[3] *Request for Information related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence*, Department of Commerce, National Institute of Standards and Technology, Docket Number 231218-0309, RIN 0693-XC135, 88 Fed. Reg. 88368 (Dec. 21, 2023) (the "RFI").

further guidance around such applications and use cases based on how deployers and users will be implementing them, rather than developing guidelines that apply generally to models without use context.

In these comments, CTA outlines factors that NIST should consider as it collects information. Because the agency will not be promulgating new rules but instead establishing guidelines and best practices for AI safety, NIST should avoid pronouncements that could lead to potential scenarios where a generally applicable guideline or practice is applied in potentially different ways by sector-specific regulators.

I. NIST's Development of a Companion Resource to the Risk Management Framework for Generative AI Should Build on the Existing Framework and Leverage a Risk-based Approach (*E.O. Mandate Sec. 4.1(a)(1)(A) & (C)*)

NIST's AI Risk Management Framework ("AI RMF") is widely regarded as a leading voluntary framework for managing risks in the AI ecosystem. CTA applauds the agency's work to create a flexible and voluntary risk management framework for managing risks unique to generative AI. Indeed, CTA's proposed National AI Policy Framework calls for all companies developing AI to adopt the AI RMF as a means of managing AI risks, together with global industry best practices for AI governance and risk management. To that end, CTA members view the RMF as a valuable tool to establish enterprise-wide processes. CTA members' experience with the AI RMF over the last twelve months has yielded several insights that may be useful for NIST as it begins to develop a companion RMF for generative AI.[4]

First, any new companion risk management framework for generative AI should continue to be presented as a voluntary framework, rather than an explicit (or implicit) mandate to engage in risk assessment protocols when developing or deploying generative AI. One reason the existing AI RMF is successful is that the voluntary nature of the framework recognizes the differing roles of AI developers and deployers, and it also provides significant flexibility to both in leveraging all, or part, of a voluntary framework in a manner best suited for their organization. The AI RMF should also be leveraged alongside other industry initiatives and practices designed to help reduce risk. Indeed, CTA has published several industry resources to assist companies in the development of a trustworthy and safe AI ecosystem, including best practices for identifying and managing bias in healthcare. NIST should ensure that the new companion RMF for generative AI is also published as a voluntary framework, which individual organizations can choose to leverage as they see fit.

Second, the utility of the AI RMF is the flexibility and variability inherent in the framework. The AI RMF is framed broadly such that procedures and protocols in the RMF can be applied and adapted to a broad variety of AI systems and applications. The companion

[4] CTA's Policy also accounts for a third category of actors: "Implementers," or end users that implement or incorporate functionality and outputs of AI Systems for their own internal, or potentially external, uses. Such actors may engage in certain activities that could increase potential risks for other persons or entities. For example, in the case of generative AI applications, end users of large language model systems can use the outputs of the model for many potential purposes and can create risks for potential victims. For high-risk applications such as these, Implementers will be required to adopt certain risk mitigation obligations and must also adhere to the duties of developers and deployers outlined in CTA's policy and regulatory framework.

resource should therefore represent an iterative development of the AI RMF specific to harms unique to generative AI, but which retains the inherent flexibility as well as the utility of the underlying framework. In other words, the companion resource should build from, and defer to, foundational principles in the AI RMF.

Third, the risk spectrum for AI systems is broad, with many systems presenting little to no effective risks, while others present potentially greater risk. The companion resource for generative AI should specifically address the risks unique to generative AI in a manner that facilitates a fulsome risk analysis, but which does not quantify and then presuppose that all measured risks are the same. It is well documented that generative AI presents certain unique risks, and NIST appropriately recognizes that such risks should be the focus of this initiative.

Finally, given the importance of global markets to CTA members, it is recommended that any companion guides and further NIST deliverables supporting the AI RMF be aligned with and contribute to relevant international standards (such as ISO/IEC). Ensuring consistency between NIST deliverables and international guidelines is important to harmonize risk management approaches and facilitate interoperable global environments for U.S. stakeholders. Additionally, this would promote further global recognition and adoption of NIST deliverables when they are reflected in international standards.

Generative AI risks can be broadly categorized and placed in two buckets: (1) those risks arising from inputs to generative AI models; and (2) those risks arising from the outputs of such models. Risks unique to generative AI arising from the input perspective include the risk of disclosure of potential confidential, proprietary, or personal information when data is input as a prompt into a generative AI model. Further, input risks include the potential that malevolent actors can use prompt-based attacks or hacks that lead the model to present outputs that could endanger individuals or society, or which are simply intentionally false, biased, and misleading. Risks arising from the outputs from these systems include the potential for generating outputs that are inaccurate, incomplete, or sometimes false (so-called "hallucinations"); these, in turn, can lead to potential outputs that may be defamatory or otherwise injurious, or potentially infringing on intellectual property rights. Other risks unique to generative AI also exist, including those pertaining to the training of LLMs. As has been well documented, risks of potential bias, discrimination, and IP infringement are acutely relevant during the training phase.

Potential generative AI risks, like others arising from "traditional" AI systems, present a range of risk profiles that may be deemed high-risk in certain cases, but not in other cases. The risks arising from inputs to generative AI models are not significantly different from the risks associated from inputs to non-generative AI models (extracting individual or group characteristics associated with the training set). Many risks may, in fact, present only moderate or limited risks, depending on the use case. For example, AI deployed in self-driving vehicles presents an inherently higher risk than generative AI systems used at home for creating shopping lists, suggesting music, controlling lights, or providing AI challengers for video-gaming. Moreover, risks for generative AI systems are unique and often differentiated from risks arising from "traditional" AI systems. The industry is already deploying unique measures to address risks unique to generative AI. The companion to the RMF for generative AI should ensure that processes for managing and governing these kinds of risks are suitably flexible to accommodate the unique challenges presented by these risks, by weighing risks against benefits. Indeed, utilizing a risk-based approach will ensure that a new companion resource for generative AI can focus solutions on the greatest potential risks, without creating overly burdensome processes on industry members leveraging the new companion RMF.

If possible, as part of the generative AI RMF, NIST should consider identifying concrete examples of the use of AI in specific industries that clearly carry higher risk. From there, NIST could develop a focused and specific multi-factor risk-based assessment that would facilitate a more nuanced approach to classifying rights-impacting AI applications on a case-by-case basis, rather than adopting a one-size-fits-all approach. NIST could then consider working with industry and sector-specific agencies to potentially incorporate these risk classifications into an overall risk management practice for companies that leverage the NIST AI RMF companion for generative AI. Notably, the EU's expected AI Act, while potentially problematic, defines "high risk" applications subject to heightened review much more narrowly.

One tool for reducing risks from generative AI model inputs and outputs is the use of model cards and/or system cards to convey relevant and important information about the model/system's development, intended uses, and potential shortcomings. CTA members continue to support transparency measures for generative AI models, and support NIST's work to increase consistency in how such model cards are deployed industry-wide to facilitate greater comparability of model cards across different deployer models. Leading AI developers have largely embraced the practice of publishing AI model cards, and some have developed specific applications to increase transparency, such as a model card "toolkit" to facilitate model card creation.[5] These developers have embraced transparency principles that facilitate the assessment of risks for the developer community, potential users, government regulators, academia, civil society, and others interested in generative AI model risks and mitigation efforts. AI model and system cards should be recognized as a key element allowing for greater transparency and assessing potential high risks arising from the development and deployment of generative AI.

II. NIST Should Proceed Cautiously in Proposing Guidelines for Evaluating and Auditing Generative AI Capabilities and Potential Harm (*E.O. Mandate Sec. 4.1(a)(1)(A) & (C)*)

NIST is seeking information from commenters to identify audit and evaluation practices or processes that can determine both AI capabilities and limitations of AI uses, with a focus on generative AI capabilities that could cause harm. CTA welcomes on-going research and efforts to develop canonical benchmarks on these issues. In particular, the development, in partnership with both public and industry stakeholders, of further NIST guidance on mechanisms for evaluating specific models of certain capabilities, such as biological or nuclear weapons knowledge, would be useful. However, before adopting or encouraging AI developers to adopt specific practices, NIST should take into account both the significant body of law that already governs the development and use of AI systems, including its own voluntary AI RMF, and the significant measures that industry leaders, trade associations such as CTA, and other governmental bodies

[5] Margaret Mitchell et al, *Model Cards for Model Reporting*, In FAT* '19: Conference on Fairness, Accountability, and Transparency (Jan. 29–31, 2019) https://doi.org/10.1145/3287560.

have already taken to encourage the development and deployment of safe and trustworthy AI, including generative AI.

First, as noted, CTA has supported efforts at the federal level to develop voluntary risk-based frameworks to address potential AI risks while enabling stakeholders to maximize the benefits of this technology, specifically supporting[6] NIST's establishment of the AI RMF.[7] Private companies, leading trade associations, government agencies, and civil society organizations are leading the way in developing additional policies, frameworks, and technical mechanisms for developing and deploying trustworthy AI, with an eye towards meeting or exceeding existing guidelines. For example, CTA published a set of its own "Guidelines for Developing Trustworthy Artificial Intelligence Systems" that are available for businesses to use when developing AI systems.[8] Industry is actively engaged in testing, assessing impacts, and developing new tools to address the potential harms that can result from the use of AI systems.

Given the increased use of these voluntary risk management frameworks and the fast-moving pace of development of this technology, we suggest that NIST proceed with caution and avoid encouraging audits or evaluations that may unnecessarily restrict or limit the deployment or use of generative AI tools and systems. Restrictions could limit the function of those systems and undermine the many benefits of generative AI available now and in the future.[9]

California and other states with consumer privacy statutes have also already proposed or enacted rules related to the use of automated tools for "profiling" that could implicate generative AI.[10] For example, the recently finalized rules implementing the Colorado Privacy Act require companies that employ profiling "for a decision that results in the provision or denial of financial or lending services, housing, insurance, education enrollment or opportunity, criminal justice, employment opportunities, health-care services, or access to essential goods or services" to provide consumers notice of:

1. the decisions that are subject to automated decision making,
2. the categories of data processed as part of the profiling,
3. a non-technical, plain language explanation of how profiling is used in the decision-making process,

[6] *See, e.g.*, Consumer Technology Association Comments, RFI - NIST AI Risk Management Framework, Docket No. 21076-01510 (filed Sept. 15, 2021), available at https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0087.pdf.

[7] On January 26, 2023, NIST released the AI Risk Management Framework (AI RMF 1.0) along with a companion NIST AI RMF Playbook, AI RMF Explainer Video, an AI RMF Roadmap, AI RMF Crosswalk, and various Perspectives.

[8] Guidelines for Developing Trustworthy Artificial Intelligence Systems (ANSI/CTA-2096), CTA, at i (Nov. 2021), available at https://shop.cta.tech/a/downloads/-/d8ec4559f62f20b9/8475bea3841a7a78/download.

[9] For the same reason, the National Security Commission on Artificial Intelligence's did not recommend regulation for AI technologies due, in part, to the "speed of technology development by the private sector . . . *See* Final Report, National Security Commission on Artificial Intelligence, at 449 (Mar. 19, 2021), *available at* https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

[10] "Generative AI shows immense promise for customer profiling and social media marketing but also raises important ethical questions around data privacy and algorithmic bias." *Digital Persona: Reflection on the Power of Generative AI for Customer Profiling in Social Media Marketing*, available at https://www.qeios.com/read/0QI028. California's "Bot Disclosure Law," prohibits the use of undeclared bots to communicate or interact with another person in California. Cal. Bus. and Prof. Code § 17940.

4. whether the system has been evaluated for fairness and accuracy,
5. the benefits and potential consequences of the decision based on profiling, and
6. information about how a consumer may choose to opt-out of such decisions.[11]

Colorado rules also provide consumers the right to opt-out of profiling in furtherance of decisions that produce legal or other "similarly significant" effects concerning a consumer, although businesses are not required to honor such requests if they employ "Human Involved Automated Processing"[12] and provide consumers with certain disclosures about the decision that incorporates the profiling process.[13] State privacy laws also ensure that consumer opt-out and access rights with regard to profiling do not extend to decisions that are only partially automated and incorporate human review within the decision-making process.

Federal sector-specific rules must also be considered before proposing guidelines that may hurt industries that are already highly regulated, such as healthcare and financial services, that use AI and generative AI systems. These industries face unique considerations that are likely to be best addressed by regulators that have developed specialized knowledge. For example, the Food and Drug Administration has been active in addressing concerns related to using automated decision making in "Software as a Medical Device;"[14] the Equal Employment Opportunity Commission has published guidance on the Americans with Disabilities Act and its impact on the use of algorithms in the hiring process;[15] the FTC has stated that existing laws already apply to the use of AI in credit eligibility decisions under the Fair Credit Reporting Act and the Equal Credit Opportunity Act;[16] the Consumer Financial Protection Bureau ("CFPB") has published guidance for financial and credit institutions who use artificial intelligence;[17] a collection of federal financial regulators including the Board of Governors of the Federal Reserve, the CFPB, the Office of Comptroller of Currency ("OCC"), and the Federal Deposit

[11] 4 CCR 904-3; 9.03(A)

[12] Defined as the automated processing of Personal Data where a human (1) engages in a meaningful consideration of available data used in the Processing or any output of the Processing and (2) has the authority to change or influence the outcome of the Processing.

[13] 4 CCR 904-3; 9.04(C). Virginia's consumer privacy law, which came into effect on January 1, 2023, also requires companies to provide consumers the ability to opt-out of profiling in furtherance of decisions that produce legal or "similarly significant" effects concerning the consumer, Va. Code Ann. § 59.1-577(A)(5), and also requires companies to conduct data protection assessments when they engage in "processing of personal data for purposes of profiling, where such profiling presents a reasonably foreseeable risk of (i) unfair or deceptive treatment of, or unlawful disparate impact on, consumers; (ii) financial, physical, or reputational injury to consumers; (iii) a physical or other intrusion upon the solitude or seclusion, or the private affairs or concerns, of consumers, where such intrusion would be offensive to a reasonable person; or (iv) other substantial injury to consumers." *Id.,* § 59.1-580(A)(3).

[14] *See* Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, Food and Drug Administration (Jan. 2021), https://www.fda.gov/media/145022/download.

[15] *See The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees*, Equal Employment Opportunity Commission (May 12, 2022), https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

[16] Andrew Smith, *Using Artificial Intelligence and Algorithms,* FTC (Apr. 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms.

[17] Consumer Financial Protection Circular 2022-03: Adverse action notification requirements in connection with credit decisions based on complex algorithms, CFPB (May 26, 2022), https://files.consumerfinance.gov/f/documents/cfpb_2022-03_circular_2022-05.pdf.

Insurance Corporation ("FDIC") have issued a Request for Information relating to Financial Institutions' Use of Artificial Intelligence, Including Machine Learning;[18] and the Department of Transportation has published a comprehensive plan on autonomous vehicles.[19]

In addition, international organizations have been actively developing guidelines with respect to trustworthy AI. For example, the ISO/IEC JTC 1 Information technology committee has engaged in a broad scope of work, developing AI best-practices addressing foundational concepts, trustworthiness aspects, data management, robustness, and cybersecurity.[20] These organizations have developed sector-specific voluntary guidelines regarding AI which have been adopted by industry participants. For example, ISO 26262 is a guideline for autonomous vehicles, and adoption of that guideline contributes to the establishment of safe and trustworthy AI systems specific to autonomous vehicles.

In the U.S., there is a growing and detrimental patchwork of complex and challenging legal and regulatory duties to address AI risk governance. To avoid confusion and the potential for conflicting obligations for companies that operate in multiple jurisdictions, NIST should ensure that any guidelines or best practices do not conflict with existing federal and state laws. This also pertains to CTA's concerns that auditing guidelines could hurt AI deployment. Although audits are not *per se* undesirable and can be an effective tool for minimizing harmful bias, inflexible and costly audits, including those that require results to be published publicly or provided to government agencies, may be overly burdensome, costly, and ineffective in eliminating actual or perceived risks. They also may create competitive concerns as other AI developers may seek to use audit results to unfairly compete with existing AI system developers.

For these reasons, any proposal endorsing the use of audits should permit operational flexibility, be properly scoped to avoid unnecessary compliance hurdles, and allow for the use of less costly or intrusive processes that may achieve the same goal – such as allowing businesses to conduct internal self-assessments and avoid publicly providing audit responses.

CTA urges NIST to instead support efforts to incentivize adoption of voluntary, risk-based approaches to AI accountability and transparency. NIST should include generative AI within the existing NIST RMF which relies on flexible risk-based assessments and recognizes the importance of proceeding deliberatively to avoid unnecessary burdens on AI development and deployment.

[18] *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning,* Request for Information and Comment, 86 Fed. Reg. 16837 (Mar. 31, 2021), https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence.

[19] Automated Vehicles Comprehensive Plan, Department of Transportation (Jan. 2021), https://www.transportation.gov/sites/dot.gov/files/2021-01/USDOT_AVCP.pdf.

[20] ISO/IEC JTC 1/SC 42 Artificial Intelligence, ISO/IEC JTC 1/ SC 27 Information security, cybersecurity and privacy protection ,

III. Industry Engagement on Red-Team Testing Should Be Leveraged in Any New Guidelines for AI Red-Team Testing (*E.O. Mandate Sec. 4.1(a)(ii)*)

NIST is directed to establish procedures to enable AI developers, in particular, developers of dual-use foundation models, to conduct red-team testing and model validation of their AI systems to find potentially harmful flaws and vulnerabilities using hacker-associated threat models to simulate how adversaries would attack an AI system. As such, red-team testing is a useful component of implementing responsible AI systems that are safe, secure, and trustworthy, and less likely to be compromised by internal or external threat actors. At the same time, NIST (and the industry) must acknowledge that while red-teaming is an important part of AI safety testing, it is still a relatively new and evolving process, and only one of several other essential evaluation and testing processes used in managing AI systems risks. Accordingly, red-teaming should be complementary to existing risk management assessments but should not be viewed as a panacea, nor should it replace existing processes that have proven effective. Instead, red-teaming should be viewed as an umbrella concept encompassing other types of safety assessments, including traditional vulnerability red-teaming, testing for unwanted output behaviors, and exploration of system capabilities. To that end, additional work identifying and defining key concepts is warranted, given the varying goals and implications of these modalities.

CTA members have extensive expertise and perspectives to leverage from their existing red-team practices for developing and deploying AI systems. A risk-based approach should be part of any guidelines, tailoring self-implemented red-teaming requirements to the different levels of risk presented by different AI systems. Because red-teaming works best when performed on a continual basis, and because it is inherently exploratory and unstructured, rigorous checklists or requirements may be counterproductive. Instead, reporting requirements for red-teaming exercises should focus on the performance of models, as contemplated in the EO, rather than mandating detailed reports on the inputs and outputs of such processes.

Companies that rely on external vendors merit a different approach. Future voluntary RMF guidance could expand on how red-teaming applies to vendor and vendor accountability, including to what extent deployers can rely on red-teaming conducted by vendors. The risks encountered are not necessarily associated with the technology itself but with its use, and that is a critical distinction in developing appropriate guidelines.

We believe there is consensus that high-risk AI systems should be subject to appropriate levels of red-team testing conducted before deployment.[21] Guidelines should emphasize that testing should be conducted by qualified experts who are independent of the product teams building those systems. Red-team testing should reflect the evolving threat landscape and include several forms of testing, such as testing threat actors' adversarial attempts to exfiltrate, extract, or corrupt an AI model's data; backdooring the model to probe security vulnerabilities; and simulating adversarial attacks against AI models and systems to find potentially harmful flaws and vulnerabilities in an AI system using hacker-associated threat models to demonstrate

[21] CTA's AI policy framework notes that system developers should inform deployers if the AI system has been tested to the extent possible for accuracy, robustness, and unintended bias, and that reasonable mitigation measures have been taken where appropriate.

how adversaries would attack an AI system. These processes can help identify system security vulnerabilities as well as the potential for generating harmful content.

CTA members include many leading technology companies that use red-team testing practices across all major categories of privacy, safety and security, and these companies also have participated in public red-teaming exercises. These practices can serve as models to leverage for the broader industry. Some companies use internal red-teams that function independently of the product teams developing the technology to test against adversarial threats. One important measure some CTA members have used is pressure testing for generative AI research and features that use LLMs with prompts that are expected to generate risky outputs. Additionally, some companies are automating red-team testing protocols and processes that apply adversarial tests to LLMs over multiple rounds. Industry practices also include developing and testing new software tools for testing and improving robustness, researching the latest attack efforts, simulating realistic adversarial activities, and sharing these practices with the broader AI research and engineering community.

IV. NIST Should Recognize the Already Significant Work on Reducing Synthetic Content Within the Industry and Leverage Best Practices Revealed in Such Efforts (*E.O. Mandate Sec. 4.5(a)*)

Another area where leading AI developers have taken significant steps in reducing risk is analyzing synthetic and potentially harmful content in both open and closed source models. First, CTA recommends a well-developed working definition of synthetic content to clarify the scope and its related risks to be incorporated in the NIST guidelines. Second, given the various limitations on existing technical approaches to labeling and identifying synthetic content, in addition to ongoing uncertainties about red-teaming practices—including red-teaming for CSAM content—NIST should provide guidance on how red-teaming can be done in a legal and safe manner. Leading AI developers are already taking significant steps on the development of technical tools and systems to address the potential harms of synthetic content. Many major AI technology companies signed on to the White House Voluntary Commitments, which include commitments to develop robust technical mechanisms to ensure that users know when content is AI generated. Similar pledges were made during the AI Safety Summit in the U.K.

In addition, leading technology companies also participate in the Partnership on AI Synthetic Media Framework which covers responsible development, creation, and sharing of synthetic media practices for industry. Most notably, many companies have updated their political content policies and verification processes for election and political advertisers, requiring disclosures when synthetic content is used in advertisements which applies to images, video, and audio content. To that end, CTA recommends that NIST guidelines should aim to recognize the existing notice and disclosure policies developers and deployers have implemented. CTA members agree that disclosures may also be appropriate when companies leverage specialized tools or agents, such as chatbots to interact with persons—in particular, in situations involving elections, politics, and public service scenarios. However, circumstances may vary, and NIST should acknowledge that mandated disclosures may not always be in the public's interest. For example, bots may facilitate the dissemination of information quickly or answer more mundane questions for consumers, where disclosures may not be necessary.

A risk-based lens to identify and distinguish potential harms from the significant benefits offered by AI is crucial in striking the right balance. Risk-based frameworks for potential new norms or practices ensure that solutions are focused on areas presenting the greatest likelihood of potential harm, while avoiding an overly broad application that could undermine beneficial uses. This balance can be achieved in part by recognizing existing content provenance initiatives, which should be embraced as potential solutions. Leveraging industry-led initiatives coupled with industry input and feedback can help avoid the potential for adopting overly prescriptive guidelines that may limit the development of useful AI tools and hinder innovation. NIST also may want to consider the types of AI offered, along with measuring the risks and benefits, in considering the impact of a generative AI framework on smaller companies that have a reduced scope of deployment.

CTA recommends leveraging existing private and public collaborations, coalitions, and consortia that are working on content provenance and authentication solutions. Similarly, public-private partnerships which include the tech industry, the public sector, NGOs, and academia, are an important component for the development and adoption of best practices for AI systems. NIST and other federal agencies should recognize these evolving efforts as effective tools and promising solutions in developing broader industry guidelines.

Many of these initiatives that exist today already address important mechanisms for developing technical guidelines for AI/ML developers, including documentations on enhancing security, assessing harms, and improving user experience. For example, CTA supports the Coalition for Content Provenance and Authenticity (C2PA), a leading industry coalition that has adopted technical guidelines for certifying the source and history (or provenance) of media content, including guidance for AI/ML developers and documentation on assessing harms, and enhancing security and user experiences. AI technology, and the authentication solutions addressing potential risks, are novel. Any new guidelines should be sufficiently flexible and open to evolving scenarios that will be identified as industry, academia, and government continue their efforts. For this reason, NIST's work should avoid prescriptive, inflexible mandates favoring particular solutions. Instead, NIST should issue guidelines that are flexible and open to new and emerging solutions, driven by existing public-private initiatives. AI can itself be used to leverage solutions in the content authentication space, and any new guidelines should recognize the utility and power of the technology to support and enable such solutions.

C2PA unifies other initiatives, including the Content Authenticity Initiative, with focus on adoption of open industry best-practices for content integrity, authenticity, and digital content provenance through open source development, cross-industry collaboration, and interoperability of technical tools. Project Origin is another coalition comprised of leading organizations from the publishing and technology realms which is working to develop a process that can confirm the provenance and technical integrity of content.

These and many other industry consortia and collaboration efforts are supported by public initiatives. In the National Defense Authorization Act of 2024, Congress recently authorized a competition to evaluate AI technology, such as tools, applications, and models for the detection and watermarking of generative AI which can be used to facilitate research, development and testing.[22] The statute also directs the Department of Defense to initiate a pilot program for open industry guidelines for embedding and authenticating digital content provenance information and metadata of publicly released official video (and audio) files.[23]

Because these initiatives are nascent, but show significant promise, NIST's recommendations should continue to support industry-led initiatives and promote R&D of technical watermarking measures through pilot programs and competitions. The power of AI presents both risks and opportunities. The opportunities are vast and far outweigh the potential risks if the technological innovations and existing industry efforts to collaborate and partner with the public sector are leveraged correctly. Also, some element of NIST's new guidelines should focus on consumer education and awareness. New tools will be more effective and credible if consumers understand how these tools work to benefit them by encouraging consumer due diligence to ensure the accuracy and validity of content presented from these AI systems.

V. U.S. Consensus Principles Should Frame Contributions to Advance Responsible Global Technical Guidelines for AI Development (*E.O. Mandate Sec. 11(b)*)

Consensus is growing around principles and guidelines for deploying safe and secure AI, but the volume of overlapping and potentially inconsistent international, national, and civil society frameworks or guidelines may impede a global consensus on baseline commitments. The volume of competing frameworks also presents a practical problem for both the public sector entities, which must leverage limited resources to support sometimes inconsistent guidelines, and the private sector entities, which must make difficult decisions about the public sector actors with which to engage –or risk losing a voice in deliberations in sometimes closed-door collaborations. This is a special challenge for startups, entrepreneurs, and small businesses, especially those deploying lower risk and higher benefit AI systems. Furthermore, competing or conflicting guidelines at a global level can create market access barriers and increased implementation burdens for U.S. industry.

U.S. companies are global leaders in the development of AI and have created such innovation in economic environments founded on free market principles under the oversight of democratic institutions. The norms, principles and authorities embraced here in the U.S., which serve as the foundation for global technology leadership, should frame U.S. positions for emerging global guidelines. CTA members also encourage further support for ISO/IEC efforts, including ongoing and intensive engagement on guidelines developed by U.S. government agencies. U.S. guidelines should be incorporated into regulatory and guidance documents to the maximum extent possible. For these reasons, NIST should develop a plan that leverages U.S. consensus principles and frameworks as a foundation for its deliverables, and that contributes these deliverables to the U.S. technical committees for international guidelines. This approach would promote the development and adoption of international guidelines with NIST counterparts in other countries, role-modeling the National Standards Strategy for Critical and Emerging

[22] Pub. Law. No. 118-31 §1543.
[23] *Id.*at §1524.

Technology (USG/ NSSCET) commitments to the World Trade Organization's Technical Barriers to Trade (WTO TBT) Agreement. [24]

VI. Conclusion

NIST plays a leading role as an agency with the technical expertise and capabilities to craft flexible, useful voluntary frameworks and guidance for this emerging sector of the economy. NIST's work will make a substantial contribution to address areas that intersect with the deployment of AI systems, including information technology, security, privacy, civil rights and civil liberties, customer experience, and workforce management. The AI ecosystem will only thrive if emerging NIST deliverables, practices and guidance remain largely voluntary, retain the flexibility and utility of the original AI RMF, and are consistent with international standards.

Respectfully submitted,

*/s/*_____
Douglas K. Johnson
Vice President, Emerging Technology Policy

*/s/*_____
Brian Markwalter
Sr. Vice President, Research & Standards

[24] The WTO TBT Agreement obligates countries to adopt international standards as the basis of technical regulations, national standards and conformity assessment procedures - to avoid unnecessary obstacles to trade. https://www.wto.org/english/docs_e/legal_e/17-tbt_e.htm.