Computer & Communications Industry Association

Open Markets. Open Systems. Open Networks.

COMMENTS OF THE COMPUTER AND COMMUNICATIONS INDUSTRY ASSOCIATION (CCIA) RE: REQUEST FOR INFORMATION (RFI) RELATED TO NIST'S ASSIGNMENTS UNDER SECTIONS 4.1, 4.5 AND 11 OF THE EXECUTIVE ORDER CONCERNING ARTIFICIAL INTELLIGENCE (SECTIONS 4.1, 4.5, AND 11)

Pursuant to the request for information published by the National Institute of Standards and Technology (NIST) in the Federal Register at 88 Fed. Reg. 88,368 (Dec. 21, 2023), the Computer & Communications Industry Association (CCIA) submits the following comments.

CCIA is an international, not-for-profit trade association representing a broad cross section of communications and technology firms. For over 50 years, CCIA has promoted open markets, open systems, and open networks.[1] CCIA members include leading artificial intelligence (AI) system designers and operators, as well as semiconductor companies whose technology is used to operate AI systems.[2]

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

In considering how best to develop guidance for ensuring AI safety and security for artificial intelligence development and deployment, CCIA urges NIST to keep in mind the fact that models do not operate in a vacuum. While CCIA's members include leading developers of AI systems, those systems may be deployed and used by a wide variety of entities. The use cases to which AI is applied are already quite varied, and the breadth of applications is only increasing each day. Because AI safety and security is heavily influenced by not just the design and development of the AI system, but also its deployment and how it is used and applied, any guidance should include standards and practices for AI deployers and users, not just developers. NIST should also consider whether there are any unique aspects to AI systems that would require additional or distinct guidance to the providers of the cloud computing systems on which many models are trained and operate.

Further, while AI is a rapidly developing area of technology, it remains an area that is built on many other technologies. Those areas of technology have existing standards that, in many cases, may already be useful to apply to AI. For example,

[1] For more, visit https://www.ccianet.org.
[2] A list of CCIA members is available at https://www.ccianet.org/about/members.

existing cybersecurity standards provide useful approaches for securing AI models and data. In addition, where international standards—such as ISO/IEC 42001—already exist, CCIA recommends NIST leverage those standards to the greatest extent possible. Any NIST guidance should use existing standards, whether developed by NIST or by international standards bodies, as a starting point where it is possible.

Finally, a number of CCIA members have either committed to their own set of principles and/or formed an agreement with the White House on a set of principles for AI safety. Those principles may be a useful source of tactics for AI risk management and evaluation. CCIA has synthesized these member principles into our *Understanding AI: A Guide To Sensible Governance* whitepaper.[3] CCIA, alongside its members, have put forward the following principles critical to responsible AI development:

- Design for social benefit.
- Design to avoid unfair outcomes.
- Analyze and minimize risks as you design.
- Consider the risks to third parties from AI systems during design, but also the benefits.
- Use up-to-date safety, security, and privacy best practices.
- Monitor and govern identified risks in deployed systems.
- Provide appropriate disclosures for deployed AI systems.

These general principles should form the backdrop to any NIST guidance, but also illustrate the approach CCIA members are already taking to AI risk management.

A Companion to the NIST Risk Management Framework

CCIA's members support the NIST Risk Management Framework (RMF) approach. Many are already applying the RMF to their AI activities, including new technologies such as generative AI. This work is done alongside existing research work on AI trust and safety. For example, CCIA member Google first proposed the use of model cards[4] to help provide increased transparency for AI models. Subsequently, they have worked to increase consistency in model card deployment and to provide tools to make it easier to document AI model information and to create model cards.

With specific respect to generative AI, CCIA believes that more research is required to develop strong evaluation capabilities. Generative AI evaluation has not yet reached a point at which good benchmarking is available. Members would welcome

[3] https://ccianet.org/wp-content/uploads/2023/06/CCIA_Understanding-AI.pdf
[4] https://arxiv.org/pdf/1810.03993.pdf

additional research and efforts to better understand how to effectively evaluate generative AI systems, whether via NIST directly or via NIST cooperating with international entities like ISO or OECD.

Further, while CCIA members have significant knowledge in many arenas of AI security and safety, there are some arenas where members lack knowledge, such as in evaluating a model's ability not to produce chemical, biological, or nuclear (CBN) weapons knowledge. A related problem is that of ensuring that generative AI systems are not used to generate child sexual abuse material (CSAM). Members stand ready to work with public and industry stakeholders to develop best practices for ensuring that generative AI systems are not used for these purposes.

AI and cybersecurity

While AI may provide avenues that "enhance[e] or otherwise affecting malign cyber actors' capabilities, such as by aiding vulnerability discovery, exploitation, or operational use", those same avenues can also be employed by security experts, for the same purposes, to better harden critical systems against attack and to better secure AI user privacy. As just one example, if an AI could automatically discover vulnerabilities, such a system could be employed pre-release by AI developers to ensure that their own systems are not subject to that type of attack.

CCIA urges NIST to consider the positive benefits of these sorts of dual-use applications, not just the potential risks, as it develops its guidance.

Red teaming

With regard to red-teaming, while it is a critical technique in AI safety testing, it is not the only technique that should be used, nor is it the only technique that members currently use. In fact, it is not even a single technique in the AI space. It includes traditional red-teaming, where the red team seeks to exploit vulnerabilities. But in the AI context the term may also include activities like testing for unwanted output behaviors via hostile prompt engineering, or simply testing for the outer reaches of system capabilities. As each of these activities may have different metrics and goals, in creating any guidance that relies on red-teaming, additional definition of what red-teaming consists of may be required to ensure that the guidance is applied consistently by all stakeholders.

Red-teaming is also an activity that may not be well-suited to an approach reliant on specific checklists or requirements. Attackers do not apply checklists, but rather unstructured exploration. As such, red-teaming guidance should tend more towards best practices, process guideposts, and general principles to ensure that it is not seen as simply one more box to check. CCIA also strongly recommends that red-teaming not be treated as a single milestone as part of development, but rather that

guidance makes clear that red-teaming should be done continually both before and after deployment.

While CCIA members have applied significant effort to safety and security against the production of content related to arenas like CBN weapons or CSAM, there are sometimes unclear legal risks to performing red-teaming in these arenas that have sometimes stymied efforts. For example, red-teaming to produce CSAM content may itself produce such illegal material. CCIA's members would appreciate guidance on how such red-teaming can be done legally and safely so that they can protect the public from these harms.

With regard to NIST's inquiry on the economic feasibility of AI red-teaming for small and large organizations and the appropriate unit of analysis for red-teaming, CCIA believes that AI red-teaming is feasible for large and small organizations alike. Red-teaming should be viewed as an integral part of AI development and resources should be dedicated to it from the start of development; an entity should never build an AI system it is incapable of red-teaming, whether by itself or via a third-party security consultant. And the appropriate unit of analysis is all aspects of AI. Red-teaming at the model/system level can address risks common across any application of the model, but specific deployments and applications may create new risks that also require analysis. As such, red-teaming should be an approach applied throughout the AI deployment chain.

Finally, because AI systems are likely to be used at different levels of one business and between separate entities, CCIA strongly suggests that NIST consider creating an AI vulnerability disclosure center, similar to NIST's existing Common Vulnerabilities and Exposures program. This will help ensure that red-teaming knowledge gained by one entity can be communicated more broadly.

Synthetic Content Risk Reduction

At the outset, CCIA notes that while its members work to avoid risks from synthetic content, they can't do it alone. Addressing these risks will require a whole-of-society approach. Technical mitigations can be, and are being, deployed by CCIA members, but those mitigations will ultimately fail unless other stakeholders are also brought into the work.

Further, while CCIA members are actively working to develop approaches to watermarking synthetic content, as well as to identify synthetic content that has not been watermarked, these are currently limited and may never be a true solution. Technical limitations to these approaches may not be overcome, though CCIA's members are working to do so. Because it is not clear whether these approaches will ultimately be feasible and usable, synthetic content guidance and

risk management cannot and must not rely solely on watermarking or detection approaches.

Synthetic content risk is not purely a technical risk, but rather a social and societal one as well. Non-technical expertise, such as social science, human psychology, language expertise, and other such backgrounds may be critical to effectively governing generative AI. Alongside technical measures, it is critical that policy interventions in other arenas, as well as user education, be used. For example, music industry stakeholders could provide provenance technology to allow consumers to be sure that the music they buy is genuinely from that artist, as it is far easier to technologically prove artist provenance than it is to provide non-removable provenance flags for AI-generated content or detect synthetic content generated by an AI system.

The same applies in the legal and regulatory realm, where the most effective mitigation may not be a technical one but rather one that solves the problem via social mechanisms. For example, synthetic content produced by an AI system being misrepresented as reality is generally a problem because of the misrepresentation, not because of the synthetic content. A video of George W. Bush singing a mashup of "Imagine" and "Walk On The Wild Side"[5] is a creative and entertaining application; an ad depicting a political opponent saying something they never said is not. Technically distinguishing between the two may be impossible, and is certainly difficult. Ultimately, legal and policy mechanisms to punish abusive user conduct may prove far more effective than technical mitigations.

Because of this, we encourage NIST, in partnership with other federal government stakeholders, to facilitate further discussion incorporating a wide variety of stakeholders including representatives of various industries, privacy advocates, security experts, and other civil society and public interest groups to discuss different roles that these varied stakeholders can play.

CCIA appreciates NIST's efforts to ensure that AI can be deployed responsibly and safely, and would be happy to further assist.

---

[5] *See* https://www.youtube.com/watch?v=GmH50pEEEY0. This video was not produced via artificial intelligence, but is a good example of the sort of creative uses to which AI might be applied.

Respectfully submitted,

Joshua Landau
Senior Counsel, Innovation Policy
Computer & Communications Industry Association
25 Massachusetts Ave NW
Suite 300C
Washington, DC 20001
jlandau@ccianet.org