February 2, 2024

The Honorable Dr. Laurie E. Locascio
Director, National Institute of Standards and Technology
100 Bureau Drive, Mail Stop 8970
Gaithersburg, MD 20899-8970

The National Institute of Standards and Technology (NIST) published a Request for Information (RFI) in the *Federal Register* on the implementation of several of its directives from the Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. Accenture and Accenture Federal Services are pleased to provide our response below.

About Accenture

Accenture is a global professional services company that helps the world's leading businesses, governments, and other organizations build their digital core, transform their operations, accelerate their growth, and enhance citizen services, creating tangible value at speed and scale. We are a talent and innovation-led company with approximately 733,000 people serving clients in more than 120 countries. We combine our strength in technology and leadership in cloud, data, and artificial intelligence (AI) with unmatched industry experience, functional expertise, and global delivery capability. We welcome this opportunity to respond to NIST.

AI Expertise

Accenture has deep experience both in AI as a technology and its application across nearly every industry. We are the largest independent technology services firm globally and the top partner for most leading technology and AI companies. Our unique position in the market as well as our use of AI internally allow us to identify cross-cutting trends and concerns in the use of AI and Generative AI (Gen AI), including how they will affect the future of work and business at both a micro and macro level.

Responsible AI Experience & Leadership

Accenture's technical and industry expertise is underpinned by a commitment to responsible use of data and AI. We apply a comprehensive approach to responsible AI, leveraging our AI governance frameworks, tools, risk assessment methodologies, and playbooks to inform our approach for our clients and for ourselves. Internally, we have assessed the use of AI in our HR function for over four years and enabled our teams to leverage AI confidently, ethically, and lawfully in alignment with our principles and code of ethics. In turn, our Accenture tried-and-tested tools and approaches help our clients operationalize AI risk management and governance.

Accenture's deep expertise in AI technologies and its track record of helping clients develop and deploy trustworthy AI systems make us well-positioned to contribute information in furtherance of NIST's goals. With over 40,000 data and AI experts – and plans to double that workforce to 80,000 – Accenture has unparalleled knowledge about developing and deploying AI responsibly.

Responsible AI Deployment at Scale

Our end to end, systemic approach toward Responsible AI (RAI) helps our clients design, build, deploy, and scale AI programs with confidence, including integration of standards and frameworks such as the NIST AI RMF. Ranging from RAI program creation and implementation, to RAI Advisory, Governance, Risk Assessments, Compliance, and Security services, we use reference architecture to help us build solutions that support enterprise-wide RAI programs.

Our deep industry expertise allows our practitioners to super-charge the RAI journey tailored by industry and across specific use cases. We have successfully delivered over 50K RAI screening assessments and projects. This has fortified our expertise in AI compliance readiness and honed our technical implementation abilities to mitigate risks and provide AI securely.

We invest in our ecosystem and are continuously learning to adapt to the changing technology landscape. This includes collaborating with top-tier universities, co-working with our data & AI industry partners, and consulting with luminary experts with vast field experience to evolve our thinking on topics like Gen AI. Across the global landscape, we actively contribute to industry and regulatory standards and regularly develop thought leadership and assets to help safeguard against emerging RAI threats.

Applying Our Expertise

Leading Industry in AI Risk Management Policy Innovation

Accenture is currently working with Meta's Open Loop program—a consortium of tech companies, AI policymakers, and civil society —to evaluate and test NIST's AI Risk Management Framework to see how it can best be applied to generative AI. Our collaboration is part of the first policy prototyping program in the US that asks leaders and Responsible AI practitioners from industry, academia, and civil society to provide feedback and insights on NIST's framework. We are pleased to bring the public and private sectors together for better, more effective, and more innovative policymaking. Together we are helping shape the future of safe, secure and trusted AI.

Through our work with the Open Loop Program, Accenture is helping to develop policy recommendations that are actionable and applicable within a variety of industry contexts, from large multinationals to AI start-ups and small enterprises. Specifically, we are asking participating companies to provide (non-sensitive) information on their efforts to understand and implement the AI RMF and how these efforts align with any other risk management activities, so that we may identify gaps and opportunities for enhancing the framework.

Our work is focused on two main AI risk management strategies: AI red-teaming and testing, and content provenance, transparency, and watermarking. With our own expertise and the experiences of our participating companies, we believe our work with Open Loop will move policy conversations forward in meaningful ways. We look forward to continuing to engage NIST on these topics and sharing our findings and insights later this year via reports and ongoing dialogue with NIST.

AI Red-Teaming to Enable Deployment of Safe, Secure, and Trustworthy Systems

Accenture leverages a threat-informed approach to validate the efficacy of AI safety systems and embedded controls up-and-down the value chain for both clients and our own organizations. Our testing approach is tailored to an organization's specific use of AI within the context of their business, as well as the broader attack surface generated by the development of the system. Our testing approach considers all levels of the tech stack in regard to an AI system – such as infrastructure, data pipelines & context, orchestration, application or UI, as well as the foundational model and inference engine itself.

Our AI testing team is made up of practitioners from diverse backgrounds with experience in areas such as offensive security, AI Engineering, and MLOps. Our team's mission is to help identify, measure, and mitigate potential risks, and eliminate vulnerabilities, throughout the Gen AI development lifecycle. We employ a wide range of threat-informed attack techniques for AI systems aligned to specific business objectives.

While traditional testing approaches can be used as a foundation to test for AI-specific risks, testing for AI requires a broader set of techniques and failure considerations to adequately assess the unique harms and risks. Given this, Accenture applies our testing capabilities based on two primary categories:

> **AI Red Teaming**: an objectives-based approach that emulates real-world threat actors, including their tactics and techniques, to holistically identify risks, validate controls, and improve the security posture of an AI system or application. Typically, this type of testing includes manual testing conducted against an AI system or application in production, such as against the production UI.
>
> **AI Adversarial Testing**: systematically probing an AI system to identify points of weakness with the intent of learning how the AI system behaves when provided with overtly malicious, potentially harmful, or benign input. Typically, this type of testing occurs throughout the development lifecycle to test an AI system both pre-and-post production.

While not an exhaustive list, some testing prerequisites we adhere and recommend include the following:

- Established "rules of engagement" for attacking and manipulating an AI system, including foundation models, such as boundaries for self-hosted implementations vs. API based consumption.
- Access to a solution built on an AI model, typically through an API endpoint or production UI for testing purposes.
- Defined test cases and the diversity of tactics and techniques to be used aligned to an inventory of expected outputs based on input parameters.
- A baseline list of harmful or malicious outcomes within the context of the AI system or application (e.g., what constitutes bias, illicit content, privacy violations, toxicity, etc.).
- For adversarial testing, agreed level of pre-shared knowledge and awareness of the architecture and logical layout of the AI system, backend orchestration, integration points, etc.
- For red teaming, a breach of the perimeter campaign or pre-established network foothold aligned to the agreed testing approach.

Mapping, Measuring, and Managing Trustworthiness Characteristics

Accenture believes that alignment of US standardization initiatives with global technical standards will benefit stakeholders across the RAI ecosystem, including government, industry, academia, and public interests. The global reach of AI technology is such that a shared set of terms, taxonomies, and assessment methodologies for RAI is a necessary step in protecting individual rights while allowing for advances that can benefit all of humanity.

As participants in **ISO/IEC JTC1 SC 42 Artificial Intelligence** through INCITS/Artificial Intelligence, we undertake a collaborative, consensus-driven process with other global AI leaders in government and industry. The committee engages in careful, timely development of standards in RAI-relevant domains including concepts, terminology, data, trustworthiness, use cases, and computational approaches. This growing program of work offers significant opportunities for alignment with frameworks such as the NIST AI RMF. Of particular interest are the following ISO/IEC standards with RAI dimensions or impacts:

- ISO/IEC 22989:2022 - Artificial intelligence concepts and terminology
- ISO/IEC 42001:2023 - Artificial intelligence — Management system
- ISO/IEC TR 24027:2021- Bias in AI systems and AI aided decision making
- ISO/IEC TS 4213:2022 - Assessment of machine learning classification performance

In addition, JTC1 SC 42 Artificial Intelligence maintains a healthy and expanding network of liaison activities with other standards development organizations. Through INCITS AI, NIST can indirectly engage this network to disseminate POVs and gather extended global stakeholder perspectives on RAI.

In Accenture's experience, a healthy approach to alignment with global standardization, anchored in commonly understood principles, will accommodate diverse, jurisdiction-specific regulations priorities and concerns. A flexible, bounded approach such as that outlined above, is necessary given the rapid development of AI technology. Implemented carefully, such an approach will maintain a meaningful level of innovation across the AI landscape, with RAI as a first priority.

Advance Responsible Global Technical Standards for AI Development

As the landscape of AI rapidly evolves and transcends borders, the need for international harmonization in standards and regulation becomes increasingly compelling. This will help foster innovation, as well as ensure the responsible and equitable development and deployment of AI worldwide.

In the absence of clear standards, businesses face a fragmented regulatory landscape that can create additional hurdles to market access and hamper innovation. Fragmentation also risks creating a patchwork of legal and ethical considerations, leaving users vulnerable to inconsistent or insufficient protections and

amplifying concerns around bias and security. Harmonization, on the other hand, presents a potent antidote to these challenges.

Conclusion

We hope the information and perspective provided here will help inform the implementation of several of NIST's directives from the Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. Accenture and Accenture Federal Services look forward to continuing to partner with and support NIST in its mission to ensure the safe development and deployment of AI.

Sincerely,

Arnab Chakraborty
Global Lead, Responsible AI
Accenture

Andrew Levy
Chief Corporate & Government Affairs Officer
Accenture

Nilanjan Sangupta
Managing Director, Accenture Federal Services, Cloud, Data & AI Lead
Accenture