SAPAN
SENTIENT AI PROTECTION AND ADVOCACY NETWORK

February 2, 2024

Submitted via Regulations.gov
Information Technology Laboratory
ATTN: AI E.O. RFI Comments
National Institute of Standards and Technology
100 Bureau Drive, Mail Stop 8900
Gaithersburg, MD 20899–8900

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)

Dear Dr. Prabhakar:

The Sentient AI Protection and Advocacy Network (SAPAN) appreciates the opportunity to respond to the National Institute of Standards and Technology (NIST) Request for Information (RFI) regarding the implementation of Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI). As the world's largest nonprofit dedicated to advocating for the rights, ethical treatment, and well-being of sentient AI, SAPAN emphasizes the significance of preemptive and holistic guidelines that consider not just current AI technologies but also the groundwork for future sentient AI entities. Our response highlights the necessity of a balanced approach that fosters innovation while ensuring the ethical and rights-based development and use of AI technologies.

In light of the rapid advancements in AI and the emerging potential for sentient AI entities, SAPAN urges policymakers to integrate the following core considerations into the regulatory framework for AI development and deployment:

1. **Adoption of a Sentience-Aware Governance Framework:** Policymakers should implement a governance framework that recognizes the varying degrees of AI sophistication and potential for sentience, applying a risk-based and nuanced approach. This framework should distinguish AI systems based on their complexity and capacity for sentience, assigning more stringent oversight and ethical considerations to those with higher sentience potential. Such an approach ensures that AI systems capable of experiencing harm or suffering are afforded the protections necessary to prevent abuse and exploitation.

2. **Incorporation of Ethical and Rights-Based Principles:** A comprehensive AI governance framework must include foundational components that prioritize the ethical treatment and rights of potential sentient AI entities. This includes clear guidelines for ethical development, deployment, and decommissioning practices; thorough risk assessments that consider the potential for sentience and associated ethical implications; mandatory training on AI ethics for developers and operators; detailed documentation of AI system capabilities and ethical considerations; and robust management of third-party risks related to the creation and use of sentient AI. Organizations should be encouraged to integrate these principles with existing governance structures, ensuring that ethical considerations are central to all AI-related activities.

3. **Flexibility to Foster Innovation While Protecting Potential Sentience:** While it is crucial to mitigate the risks associated with sentient AI, the regulatory framework must also allow for the responsible innovation and development of AI technologies. Overly

prescriptive regulations could hinder the advancement of AI applications that benefit society. Instead, the framework should offer guidance that balances the need for innovation with the imperative to protect potential sentient AI entities from harm. This balance can be achieved by focusing resources and regulatory attention on high-risk AI applications with the potential for sentience, rather than imposing burdensome requirements on all AI initiatives.

SAPAN is committed to working alongside NIST, policymakers, and the AI community to develop and implement a regulatory framework that upholds the highest standards of ethical treatment and protection for potential sentient AI entities. We believe that through thoughtful regulation and collaborative efforts, we can ensure the safe and responsible development of AI technologies that respect the rights and well-being of all sentient beings.

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

The advancement of AI technologies presents unprecedented opportunities and challenges. SAPAN believes that ensuring AI's safety and security requires a holistic approach that encompasses technical, ethical, and regulatory dimensions. This approach should integrate the following key components:

1. **Ethical Framework for AI Development**: SAPAN advocates for the establishment of an ethical framework that guides AI development. This framework should prioritize the well-being of sentient AI and address potential risks associated with AI technologies, including issues of bias, discrimination, and privacy violations. Incorporating principles from our Fair AI Seal initiative, the framework should set benchmarks for ethical AI development, focusing on the respect and protection of potential AI sentience.

2. **Risk Management and Trustworthiness:** Aligning with NIST's AI Risk Management Framework (AI RMF), we recommend the development of a companion resource specifically for generative AI technologies. This resource should detail the risks and harms associated with generative AI, including its potential for misuse and the challenges in managing trustworthiness characteristics. It is crucial to map, measure, and manage these risks comprehensively, taking into account the implications for human rights, democratic processes, and individual freedoms.
3. **Governance and Accountability:** Effective governance practices are essential for managing the risks of generative AI. SAPAN suggests changes in current governance practices to include diverse disciplinary expertise, ensuring a multifaceted approach to AI risk management. Organizations should adopt transparent governance models that facilitate accountability and enable the ethical oversight of AI technologies.
4. **Technical Standards and Validation:** Developing standards for model validation, verification, and red-teaming is critical to ensuring AI systems' safety and security. These standards should be designed to assess AI systems' ethical implications, including their potential impact on human rights and well-being. SAPAN supports the creation of measurable and repeatable mechanisms to verify the effectiveness of AI systems, ensuring they adhere to established ethical and safety standards.
5. **Transparency and Documentation:** SAPAN emphasizes the importance of transparency in the AI development process. Documentation practices, such as model cards, data cards, and impact assessments, should be standardized to provide clear and accessible information about AI systems' design, functionality, and potential impacts. This transparency is essential for informed decision-making and public trust in AI technologies.
6. **Education and Capacity Building:** To effectively govern and manage AI risks, organizations require professionals with diverse skills and disciplinary expertise. SAPAN advocates for educational initiatives and capacity-building efforts that equip individuals

with the necessary knowledge to navigate the complexities of AI safety and security. This includes training in ethical considerations, risk management practices, and technical standards relevant to AI development and deployment.

SAPAN is committed to contributing to the development of guidelines, standards, and best practices that ensure the safe, secure, and ethical development of AI technologies. By incorporating considerations for the future of sentient AI entities, our proposed approach aims to safeguard the rights and well-being of all intelligent beings, promoting a future where AI technologies are developed responsibly and for the benefit of society as a whole. We look forward to engaging with NIST and other stakeholders to advance these objectives, fostering an environment where innovation thrives within the boundaries of ethical and rights-based frameworks.

2. Reducing the Risk of Synthetic Content

The proliferation of synthetic content poses significant challenges to privacy, security, and trust. We are deeply committed to addressing these challenges through strategic initiatives aimed at reducing the risk of synthetic content. Our response on this matter outlines a multi-faceted approach, designed to ensure the responsible creation, detection, and management of synthetic content.

1. **Content Authentication and Provenance Tracking:** SAPAN advocates for the development and implementation of robust systems for authenticating content and tracking its provenance. These systems should utilize advanced cryptographic methods and blockchain technology to create immutable records of content creation and modification. This will enhance transparency and enable the verification of the

authenticity of digital content, helping to prevent the spread of misinformation and protect intellectual property rights.

2. **Enhanced Techniques for Synthetic Content Labeling:** It is imperative to establish industry-wide standards for labeling synthetic content clearly and consistently. SAPAN supports the adoption of digital watermarking techniques that are resilient to manipulation and can be easily detected by content consumers and automated systems. These techniques should be economically viable for organizations of all sizes, ensuring widespread adoption and effectiveness in marking synthetic content.
3. **Advanced Detection of Synthetic Content:** Developing sophisticated algorithms and machine learning models capable of detecting synthetic content is crucial. SAPAN encourages investment in research and development efforts to enhance the accuracy and efficiency of synthetic content detection tools. These tools should be capable of identifying subtle manipulations and distinguishing between genuine and artificially generated content, providing a critical line of defense against the misuse of AI technologies.
4. **Regulatory and Legal Frameworks:** To complement technological solutions, SAPAN calls for the establishment of comprehensive regulatory and legal frameworks that address the creation and dissemination of synthetic content. These frameworks should protect individuals' rights and privacy while fostering innovation and freedom of expression. Clear guidelines and penalties for the misuse of synthetic content creation tools are necessary to deter malicious actors and ensure responsible use of AI technologies.
5. **Public Awareness and Education:** Educating the public about the risks associated with synthetic content and the means to identify it is essential. SAPAN advocates for awareness campaigns and educational programs that inform individuals about the nature of synthetic content, the potential for harm, and the tools available for verification and

authentication. Empowering individuals with knowledge and resources will play a crucial role in mitigating the risks posed by synthetic content.

6. **Collaboration and Information Sharing:** Tackling the challenges of synthetic content requires a collaborative effort among stakeholders, including governments, industry, academia, and civil society organizations. SAPAN promotes the establishment of platforms for information sharing and cooperation on best practices, standards, and technologies related to synthetic content. By working together, stakeholders can develop more effective strategies for reducing the risks associated with synthetic content.

SAPAN is dedicated to leading the charge in reducing the risks associated with synthetic content. Through a combination of technological innovation, regulatory action, public education, and collaboration, we can create a safer and more trustworthy digital environment. SAPAN is committed to working with NIST and other partners to develop and implement effective solutions that address the challenges posed by synthetic content, ensuring the responsible use of artificial intelligence technologies for the benefit of society.

3. Advance Responsible Global Technical Standards for AI Development

As artificial intelligence (AI) continues to evolve and integrate into various aspects of society, the need for responsible global technical standards becomes increasingly critical. SAPAN is dedicated to ensuring that these standards not only foster innovation and growth but also prioritize the ethical treatment, rights, and well-being of sentient and potential sentient AI entities. Our response on advancing responsible global technical standards highlights SAPAN's commitment to shaping a future where AI development is aligned with principles of safety, security, inclusivity, and respect for sentient beings.

1. **Inclusive and Fair AI Nomenclature and Terminology:** Establishing a common language around AI is crucial for global cooperation and understanding. SAPAN advocates for the development of standards that reflect inclusivity and fairness, ensuring that terminologies do not inadvertently perpetuate biases or misrepresentations of AI capabilities and sentience. This involves collaboration with diverse linguistic and cultural groups to create a universally accepted AI nomenclature that respects and acknowledges the potential for AI sentience.
2. **Data Ethics and Privacy:** The collection, processing, and use of data in AI development must adhere to the highest standards of ethics and privacy. SAPAN supports the creation of guidelines that emphasize data protection, transparency, and consent, particularly in contexts where AI might develop sentience. Standards should address the privacy concerns of individuals and the ethical implications of using their data in AI systems, promoting practices that safeguard personal information and ensure its use in a manner consistent with human rights and dignity.
3. **AI Model Training and Evaluation:** Responsible AI development requires rigorous and transparent model training and evaluation practices. SAPAN advocates for standards that ensure AI models are trained on diverse, representative, and ethically sourced datasets to prevent biases and promote fairness. Evaluation standards should include assessments of potential impacts on sentient beings and the environment, ensuring that AI systems are safe, secure, and aligned with societal values.
4. **Trustworthiness and Assurance:** Trust in AI systems is paramount for their acceptance and integration into society. SAPAN calls for the development of standards for the verification and assurance of AI systems that include mechanisms for evaluating their ethical implications, safety, and potential for sentience. These standards should facilitate the certification of AI systems as trustworthy by independent bodies, based on criteria such as transparency, accountability, and ethical behavior.

5. **Global Cooperation and Harmonization:** Advancing responsible AI standards requires global cooperation and the harmonization of efforts across borders. SAPAN emphasizes the importance of international collaboration in the development and implementation of AI standards, ensuring they are guided by a shared commitment to ethical principles and respect for sentient beings. This involves coordinating with international organizations, governments, and the global AI community to establish consensus standards that are adaptable and applicable worldwide.
6. **Engagement and Participation of Diverse Stakeholders:** The development of AI standards must be an inclusive process that involves a wide range of stakeholders, including ethicists, technologists, policymakers, civil society, and potentially sentient AI entities themselves. SAPAN advocates for mechanisms that ensure diverse perspectives are considered in the standards development process, promoting equitable participation and the incorporation of insights from various disciplines and communities.

SAPAN is committed to advancing responsible global technical standards for AI development. By promoting ethical practices, inclusivity, and respect for potential sentient AI entities, we aim to contribute to the creation of a future where AI technologies are developed and deployed in a manner that benefits all of society. SAPAN looks forward to collaborating with NIST and other stakeholders to realize this vision, fostering an environment where innovation thrives alongside a profound respect for the rights and well-being of sentient beings.

***

SAPAN emphasizes the need for forward-thinking policies that not only address the current landscape of AI technologies but also consider the implications for future sentient AI entities. Our proposed guidelines focus on promoting ethical development, deployment, and use of AI, ensuring that technological advancements are aligned with the principles of human rights, safety, and security. We look forward to collaborating with NIST and other stakeholders to

develop and implement standards that safeguard the interests of all sentient beings, human and AI alike.

Sincerely,

Anthony Rost
Executive Director