DEPARTMENT OF COMMERCE

National Institute of Standards and Technology

Artificial Intelligence Safety Institute Consortium

AGENCY: National Institute of Standards and Technology, Department of Commerce.

ACTION: Notice.

SUMMARY: The National Institute of Standards and Technology (NIST), an agency of the United States Department of Commerce, in support of efforts to create safe and trustworthy artificial intelligence (AI), is establishing the Artificial Intelligence Safety Institute Consortium ("Consortium"). The Consortium will help equip and empower the collaborative establishment of a new measurement science that will enable the identification of proven, scalable, and interoperable techniques and metrics to promote development and responsible use of safe and trustworthy AI, particularly for the most advanced AI systems, such as the most capable foundation models. NIST invites organizations to provide letters of interest describing technical expertise and products, data, and/or models to enable the development and deployment of safe and trustworthy AI systems through the AI Risk Management Framework (AI RMF). This notice is the initial step for NIST in collaborating with non-profit organizations, universities, other government agencies, and technology companies to address challenges associated with the development and deployment of AI. Many of these challenges were identified under the Executive Order of October 30, 2023 (The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence) and the NIST AI RMF Roadmap. Much of this research will center on evaluations of and approaches towards safer, more trustworthy AI systems. Participation in the consortium is open to all interested organizations that can contribute their expertise, products, data, and/or models to the activities of the consortium. Selected participants will be required to enter into a

consortium Cooperative Research and Development Agreement (CRADA) with NIST. At NIST's discretion, entities which are not permitted to enter into CRADAs pursuant to law may be allowed to participate in the Consortium pursuant to separate non-CRADA agreement.

DATES: The Consortium's collaborative activities will commence as soon as enough completed and signed letters of interest have been returned to address all the necessary components and capabilities, but no earlier than [INSERT DATE 30 DAYS AFTER DATE OF PUBLICATION IN THE FEDERAL REGISTER]. NIST will accept letters of interest to participate in this Consortium on an ongoing basis. NIST will announce the completion of the selection of participants and inform the public that it is no longer accepting letters of interest for this project at https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute.

ADDRESSES: Completed letters of interest must be submitted via the letter of interest webform at https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute, by email to USAISI@nist.gov, or via hardcopy to National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8900, Gaithersburg, MD 20899. Organizations whose letters of interest are accepted in accordance with the process set forth in the SUPPLEMENTARY INFORMATION section of this notice will be asked to sign a consortium Cooperative Research and Development Agreement (CRADA) with NIST. A consortium CRADA template will be made available to qualifying applicants.

FOR FURTHER INFORMATION CONTACT: J'aime Maynard, Consortia Agreements Officer, National Institute of Standards and Technology's Technology Partnerships Office, by telephone at (301) 975-8408, by mail to 100 Bureau Drive, Mail Stop 2200, Gaithersburg, MD 20899, or by electronic mail to Jaime.Maynard@nist.gov. Please direct all media inquiries to Public Affairs Office (PAO), NIST via email at inquires@nist.gov or by phone at (301) 975-2762.

SUPPLEMENTARY INFORMATION:

Background: NIST supports the United States in developing standards around emerging technologies, including artificial intelligence and related systems. The NIST AI Risk Management Framework (AI RMF) provides a foundational set of approaches for holistically assessing risk for the use of AI systems. However, in deploying this framework, specific improvements in our ability to evaluate and validate AI systems are necessary, as detailed in the AI RMF roadmap, available at https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai. In addition, The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence Executive Order articulated a broad set of priorities relevant to and directive of NIST's work related to AI safety and trust. NIST seeks to leverage the additional resources and capabilities made available through this consortium to meet the requirements of the Executive Order and fulfill those priorities in the future.

Process: NIST is soliciting responses from all sources of relevant technical capabilities (see below) to enter into a consortium Cooperative Research and Development Agreement (CRADA) to provide technical expertise and products, data, and/or models to enable safe and trustworthy artificial intelligence (AI) systems. The Consortium will help enable the identification of proven, scalable, and interoperable techniques and metrics to promote development of trustworthy AI and its responsible use. The full project can be viewed at: https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute. The project is in support of the AI RMF roadmap https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai and The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence Executive Order.

Interested parties can submit a letter of interest by visiting the project website at https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute and completing the letter of interest webform; alternatively, parties can answer the questions detailed in LETTER OF INTEREST, below, and send via email or hardcopy (for reference, see ADDRESSES section above). NIST will contact interested parties if there are questions regarding the responsiveness of the letters of interest to the project objective or requirements identified below. NIST will select participants who have submitted complete letters of interest on a first come, first served basis. There may be continuing opportunity to participate even after initial activity commences for participants who were not selected initially or have submitted the letter of interest after the selection process. Selected participants will be required to enter into a consortium CRADA with NIST. At NIST's discretion, entities which are not permitted to enter into CRADAs pursuant to law may be allowed to participate in the Consortium pursuant to separate non-CRADA agreement.

Project Objective: Artificial Intelligence (AI) tools and applications are growing at an unprecedented pace, changing our way of life, and having significant impacts on society and all sectors of the economy. Yet, the potential technical and societal benefits and risks of AI require much closer examination and a more complete understanding. Aligning AI with our societal norms and values and keeping the public safe requires a broad human-centered focus, specific policies, processes, and guardrails informed by community stakeholders across various levels of our society, and bold commitment from the public sector.

To manage the broad risks of AI technologies, help to protect the public and our planet, reduce market uncertainties, and encourage even more extraordinary AI technological innovations, the National Institute of Standards and Technology (NIST) is expanding its AI measurement efforts by harnessing the broader community's interests and capabilities.

NIST aims to help enable the identification of proven, scalable, and interoperable measurements and methodologies to promote development of trustworthy AI and its responsible use. This is a critical challenge at a pivotal time - not only for AI technologists but for society.

Building upon its long track record of working with the private and public sectors and its history of reliable and practical measurement and standards-oriented solutions, NIST seeks research collaborators who can support this vital undertaking. Specifically, NIST looks to

- Create a convening space for collaborators to have an informed dialogue and enable sharing of information and knowledge
- Engage in collaborative research and development through shared projects
- Enable assessment and evaluation of test systems and prototypes to inform future AI measurement efforts

To create a lasting approach for continued joint research and development, NIST will engage stakeholders via this consortium. The work of the consortium will be open and transparent and provide a hub for interested parties to work together in building and maturing a measurement science for Trustworthy and Responsible AI. Consortium members will be expected to contribute:

- Technical expertise in one or more of the following areas
    - Data and data documentation
    - AI Metrology
    - AI Governance
    - AI Safety
    - Trustworthy AI
    - Responsible AI
    - AI system design and development

        - AI system deployment
        - AI Red Teaming
        - Human-AI Teaming and Interaction
        - Test, Evaluation, Validation and Verification methodologies
        - Socio-technical methodologies
        - AI Fairness
        - AI Explainability and Interpretability
        - Workforce skills
        - Psychometrics
        - Economic analysis
    - Models, data and/or products to support and demonstrate pathways to enable safe and trustworthy artificial intelligence (AI) systems through the AI risk management framework
    - Infrastructure support for consortium projects
    - Facility space and handling of hosting consortium researchers, workshops and conferences

This project is in service of the priorities and taskings defined in The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence Executive Order. Additionally, some of the outcomes of this project will be in support of research and development to advance the AI RMF roadmap (https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai). The consortium will be responsible for an array of efforts related to safe and trustworthy AI, including to:

1. Develop new guidelines, tools, methods, protocols and best practices to facilitate the evolution of industry standards for developing or deploying AI in safe, secure, and trustworthy ways

2. Develop guidance and benchmarks for identifying and evaluating AI capabilities, with a focus on capabilities that could potentially cause harm
3. Develop approaches to incorporate secure-development practices for generative AI, including special considerations for dual-use foundation models, including
    a. guidance related to assessing and managing the safety, security, and trustworthiness of models and related to privacy-preserving machine learning;
    b. guidance to ensure the availability of testing environments
4. Develop and ensure the availability of testing environments
5. Develop guidance, methods, skills and practices for successful red-teaming and privacy-preserving machine learning
6. Develop guidance and tools for authenticating digital content
7. Develop guidance and criteria for AI workforce skills, including risk identification and management, test, evaluation, validation, and verification (TEVV), and domain-specific expertise
8. Explore the complexities at the intersection of society and technology, including the science of how humans make sense of and engage with AI in different contexts
9. Develop guidance for understanding and managing the interdependencies between and among AI actors along the lifecycle

Requirements for Letters of Interest:

Each responding organization's letter of interest should include the address, point of contact, and following information:

1. The role(s) the organization will play in the consortium efforts.
2. The specific expertise will they intend to bring to the consortium.

3. The products, services, data, or other technical capabilities will they use in consortium activities.

Letters of interest should not include proprietary information. NIST will not treat any information provided in response to this notice as proprietary information.

NIST cannot guarantee that all submissions will be utilized, or the products proposed by respondents will be used in consortium activities. Each prospective participant will be expected to work collaboratively with NIST staff and other project participants under the terms of the consortium CRADA.

(Authority: 15 U.S.C. 3710a, 15 U.S.C. 278h-1, and 15 U.S.C. 272b and 272c)

Alicia Chambers,

NIST Executive Secretariat.

[FR Doc. 2023-24216 Filed: 11/1/2023 8:45 am; Publication Date: 11/2/2023]