FRANK D. LUCAS, Oklahoma
CHAIRMAN

ZOE LOFGREN, California
RANKING MEMBER

Congress of the United States
House of Representatives
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
2321 Rayburn House Office Building
Washington, DC 20515–6301
(202) 225–6371
www.science.house.gov

December 14, 2023

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

Dear Director Locascio,

We are writing in regard to the establishment of the Artificial Intelligence Safety Institute (AISI) at the National Institute of Standards and Technology (NIST). Specifically, we are concerned about how the AISI will fund outside organizations and the transparency of those awards.

We applaud the NIST for its important efforts to guide the responsible development and deployment of trustworthy artificial intelligence. Developing meaningful governance for AI systems requires quantitative metrics, test methods, and accountability tools – a significant scientific challenge that the U.S. government and the broader AI community needs to get right. This challenge will require a sustained and well-resourced effort by the public sector, diverse industry stakeholders, and academia on AI research, standards development, and evaluations for AI use cases.

NIST is rightly viewed as a leader in developing a robust, scientifically grounded framework for the field of AI trust and safety research. NIST's work to establish a voluntary risk framework and related guidance for organizations creating and using artificial intelligence will be foundational to governance of these systems. The activities assigned to NIST in Executive Order 14110, "*Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,"* are a step forward in actualizing standards and evaluations of AI systems and promoting their development and adoption.[1] Specifically, the E.O. directs NIST to establish the AISI, to bring together the stakeholders required to overcome AI-related challenges.

Unfortunately, the current state of the AI safety research field creates challenges for NIST as it navigates its leadership role on the issue. Findings within the community are often self-referential and lack the quality that comes from revision in response to critiques by subject matter experts. There is also significant disagreement within the AI safety field of scope, taxonomies, and definitions.[2]

The examples are numerous. One analysis that found 'emergent capabilities' and 'sparks of artificial general intelligence' within large language models[3] was debunked by rigorous statistical analysis.[4] Some argue that artificial general intelligence is already here[5], but that conclusion is called into question by the fact that AI systems as deployed often fail to function.[6] Organizations routinely point to significant speculative benefits or risks of AI systems but fail to provide evidence of their claims,[7] produce nonreproducible research,[8] hide behind secrecy,[9] use evaluation methods that lack construct validity,[10] or cite research that has failed to go through robust review processes, such as academic peer review.[11] Some of this is certainly due to a lack of consensus in the field. However, while debate within the maturing field is good, standards and evaluations that result from NIST's work must be fit for purpose and cannot be based on speculation or methodologically unsound research.

We have learned that NIST intends to make grants or awards through the AISI to outside organizations for extramural research. There does not appear to be any publicly available information about the process for these awards—no notice of funding opportunity, announced competition, or public posting. These awards were not discussed during the public listening session on the AISI that NIST held on Friday, November 17th or the recent Congressional Staff briefing on Monday, December 11th. The process for these awards differs significantly from the information that NIST has provided to organizations interested in entering into a Cooperative Research and Development Agreement (CRADA) to participate in the consortium with NIST.[12]

Members of the House Committee on Science, Space, and Technology have long supported NIST's important work to advance trustworthiness in AI systems, including through the *National Artificial Intelligence Initiative Act of 2020* and the *CHIPS and Science* legislation. We believe this work should not be rushed at the expense of doing it right. Developing novel evaluation suites complete with appropriate metrics for AI trustworthiness across successive generations of large language models could itself take years – without taking into account how these AI systems are deployed across sectors and use cases. As NIST prepares to fund extramural research on AI safety, scientific merit and transparency must remain a paramount consideration. In implementing the AISI, we expect NIST to hold the recipients of federal research funding for AI safety research to the same rigorous guidelines of scientific and methodological quality that characterize the broader federal research enterprise.

We request a staff briefing from NIST to discuss the AISI process and use of funds. These important activities are an opportunity to mature the field of AI safety. I appreciate your dedication to addressing these important issues and look forward to continuing to work with you. If you have any questions, please contact Anna Ferrara of the Committee's Majority staff at (202) 225-6371 or Alan McQuinn of the Committee's Minority staff at (202) 225-6375.

Sincerely,

Frank D Lucas

Frank Lucas
Chairman
House Committee on
Science, Space, and Technology

Zoe Lofgren

Zoe Lofgren
Ranking Member
House Committee on
Science, Space, and Technology

Mike Collins
Chairman
Subcommittee on Research and Technology
House Committee on
Science, Space, and Technology

Haley Stevens
Ranking Member
Subcommittee on Research and Technology
House Committee on
Science, Space, and Technology

Jay Obernolte
Chairman
Subcommittee on Investigations and Oversight
House Committee on
Science, Space, and Technology

Valerie F Foushee

Valerie Foushee
Ranking Member
Subcommittee on Investigations and Oversight
House Committee on
Science, Space, and Technology

cc:

The Honorable Gina Raimondo, Secretary of Commerce

The Honorable Arti Prabhakar, OSTP Director

[1] "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," Executive Office of the President, *Federal Register*, October 30, 2023, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence;

[2] Inioluwa Deborah Raji and Roel Dobbe, "Concrete problems in AI Safety, revisited," In ICLR workshop on ML in the real world, 2020, https://drive.google.com/file/d/1Re_yQDNFuejoqjZloTgQpILosDGtt5ei/view.

[3] Sebastien Bubeck et. al, "Sparks of Artificial General Intelligence: Early experiments with GPT-4," arxiv, revised April 13, 2023, https://arxiv.org/abs/2303.12712.

[4] Ryan Schaeffer, Brando Miranda, Sanmi Koyejo, "Are Emergent Abilities of Large Language Models a Mirage?" arxiv, revised May 22, 2023, https://arxiv.org/abs/2304.15004.

[5] Blaise Aguera y Arcas and Peter Norvig, "Artificial General Intelligence Is Already Here," October 10, 2023, https://www.noemamag.com/artificial-general-intelligence-is-already-here/.

[6] Inioluwa Deborah Raji et. al., "The Fallacy of AI Functionality," FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, June 2022, p 959-972, https://dl.acm.org/doi/10.1145/3531146.3533158.

[7] "Statement on AI Risk," Center for AI Safety, 2023, https://www.safe.ai/statement-on-ai-risk.

[8] O.E. Gundersen, "Improving reproducibility of artificial intelligence research to increase trust and productivity," OECD iLibrary, 2023, https://www.oecd-ilibrary.org/sites/3f57323a-en/index.html?itemId=/content/component/3f57323a-en.

[9] Meredith Whittaker, "The steep cost of capture," interactions 28, 6, November - December 2021, 50–55. https://doi.org/10.1145/3488666.

[10] Inioluwa Deborah Raji et al., "AI and the Everything in the Whole Wide World Benchmark," NeurIPS 2021 Benchmarks and Datasets track, November 26, 2021, https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/084b6fbb10729ed4da8c3d3f5a3ae7c9-Paper-round2.pdf.

[11] For example, see Christopher Mouton, Caleb Lucas, and Ella Guest, "The Operational Risks of AI in Large-Scale Biological Attacks," RAND Corporation, 2023, https://doi.org/10.7249/RRA2977-1.

[12] "Artificial Intelligence Safety Institute Consortium," National Institute of Standards and Technology, *Federal Register,* November 2, 2023, https://www.federalregister.gov/documents/2023/11/02/2023-24216/artificial-intelligence-safety-institute-consortium.