February 1, 2024

National Institute of Standards and Technology (NIST)

100 Bureau Drive, Stop 2000

Gaithersburg, MD 20899-8900

ATTN: NIST AI Executive Order RFI Comments

RE: Request for Information in Support of Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI)

Greetings NIST,

Thank you for the opportunity to contribute to the deployment of safe, secure, and trustworthy AI systems. Recommendations have been outlined in the subsequent pages.

Sincerely,

Thomas Byrd

Making Artificial Intelligence Transparent and Trustworthy

This document will address measures necessary for building comprehensive transparent and trustworthy Artificial Intelligence (AI) systems thus reducing the risk of danger and promoting safety.

In 2023, the Biden-Harris Administration secured voluntary commitments from leading AI companies to manage the risks posed by these novel systems. This commitment included the public reporting of the capabilities, limitations, and appropriate and inappropriate use of their AI systems.[[1]](#footnote-1) The companies also committed to developing robust technical mechanisms to ensure that users know when content is AI-generated.[[2]](#footnote-2) Since the risks posed by AI are unique, the disclosure of when and how AI is used to generate outcomes, decisions, and content is imperative for public safety and equity.

The National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (AI RMF 1.0) provides guidance that increases the trustworthiness of AI systems.[[3]](#footnote-3) For example, the AI RMF 1.0 mentions that AI systems may be trained on data that can change over time sometimes significantly and unexpectedly, affecting system functioning in ways that are hard to understand.[[4]](#footnote-4) AI systems and the contexts in which they are deployed are complex, making it difficult to detect and respond to failures when they occur.[[5]](#footnote-5)

The reporting of AI system capabilities, a disclosure of the presence of AI-generated content, plus insight into the technology supply chain including security practices are key to creating trustworthy AI systems. Disclosure is necessary to account for the unique challenges associated with AI and to enable businesses and individuals to assess the compressive capabilities and risks inherent in an AI system throughout the reasonably foreseeable period of use. Since this level of granularity has not been traditionally disclosed, any proactive requirements concerning AI system transparency should require legislative action.

Statutory requirements and obligations will serve as the necessary framework to achieve consistent transparency throughout the technology lifecycle. Legislation should be enacted to create requirements in lieu of relying upon the voluntary commitments. The legislation should require manufacturers of AI systems to include with their systems a **“***transparency report.****”*** This report should require the disclosure of the scope in which AI technologies are embedded in products and the service and maintenance required to sustain the capability of an AI system and any factors that have an impact on security and privacy.

To reduce risk and promote transparency, manufacturers of AI systems should disclose information regarding training and supply chain data. Since their activities may affect the safety properties of an AI system, disclosure creates an awareness of the potential for biases and inaccuracies in outputs. Accordingly, information regarding the characteristics of the AI system should include the criteria used to assess data quality, system performance, evaluation and validation results, deployment monitoring, and setting guidelines on product safety should be part of this disclosure. Any AI system that could pose risks that are not reasonably obvious should be labeled with a warning and instructions for use.

AI systems and the contexts in which they are deployed are complex, making it difficult to detect and respond to failures when they occur. When an automaker promotes a vehicle design that incorporates state -of -the- art technology to enhance safety, the manufacturer develops numerous classifications of driver assistance technology with functional limitations. To operate the vehicle in a safe manner, the driver needs to understand that regardless of the name that the manufacturer uses to market the technology, functional limitations do not make the vehicle fully autonomous. Driver supervision is still required. Investigations by transportation safety boards highlight concerns regarding trust and transparency about the technology. Any over-reliance, lack of consumer understanding of the functionality and technology limitations were listed in conclusions by the safety boards as problematic.

Special knowledge of the AI systems illustrates the necessity for implementing requirements to defend critical networks and infrastructure. This may deter misfeasance. The need for trust and transparency resonates with AI applications across all infrastructure sectors whether transportation, defense or space exploration. In the scenario of autonomous vehicles, information regarding the characteristics of the AI system should include the criteria used to assess data quality, performance, monitoring, and guidelines to promote product safety. The implications of wayward AI systems could be calamitous. Any product that could pose risks that are not reasonably obvious should be labeled with a warning and instructions for use.

In closing, it is imperative to align the measures with a legislative framework to ensure AI systems are trustworthy and transparent. This a necessary approach to managing and mitigating risks of AI technologies and promoting responsible business practices and public accountability.

1. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/#:~:text=As%20part%20of%20this%20commitment,help%20move%20toward%20safe%2C%20secure%2C> [↑](#footnote-ref-1)
2. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/#:~:text=As%20part%20of%20this%20commitment,help%20move%20toward%20safe%2C%20secure%2C> [↑](#footnote-ref-2)
3. The National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://doi.org/10.6028/NIST.AI.100-1> [↑](#footnote-ref-3)
4. The National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://doi.org/10.6028/NIST.AI.100-1> [↑](#footnote-ref-4)
5. The National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://doi.org/10.6028/NIST.AI.100-1> [↑](#footnote-ref-5)