TECHNET
THE VOICE OF THE
INNOVATION ECONOMY

1420 New York Avenue NW, Suite 825
Washington, D.C. 20005
www.technet.org | @TechNetUpdate

September 9, 2024

National Institute of Standards and Technology
100 Bureau Drive, Mail Stop 8900
Gaithersburg, MD 20899

RE: NIST AI 800-1, Managing Misuse Risk for Dual-Use Foundation Models

To Whom It May Concern:

TechNet appreciates the opportunity to comment on the National Institute of Standards and Technology's (NIST) public draft guidance on Managing Misuse Risk for Dual-Use Foundation Models (NIST AI 800-1). Many of our nation's leading AI developers, deployers, researchers, and users are TechNet members.

TechNet is the national, bipartisan network of technology CEOs and senior executives that promotes the growth of the innovation economy by advocating a targeted policy agenda at the federal and 50-state level. Our membership includes dynamic American businesses ranging from startups to the most iconic companies on the planet and represents over 4.4 million employees and countless customers in the fields of information technology, artificial intelligence, e-commerce, the sharing and gig economies, advanced energy, transportation, cybersecurity, venture capital, and finance.

TechNet strongly values NIST's ongoing work to partner with industry, academia, and civil society to support the development of resilient and responsible standards for testing AI systems. Several TechNet members are participants in the AI Safety Consortium, providing expert perspectives and cutting-edge resources for standards development. TechNet has also previously engaged in several of NIST's feedback opportunities, submitting comments on NIST AI 600-1: AI Risk Management Framework - Generative AI Profile, NIST AI 100-5: A Plan for Global Engagement on AI Standards, and on NIST's assignments under sections 4.1, 4.5, and 11 of the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO). We appreciate the agency's consideration of our recommendations and perspectives for the finalization of NIST AI 800-1.

Consistency with the AI Executive Order
Under the Executive Order concerning AI, NIST is charged in Subsection 4.1(a)(ii) to "Establish appropriate guidelines (except for AI used as a component of a national security system), including appropriate procedures and processes, to enable developers of AI, especially of dual-use foundation models, to conduct AI

red-teaming tests to enable deployment of safe, secure, and trustworthy systems."[1] NIST cites that NIST AI 800-1 fulfills the requirements of this responsibility, however this proposed guidance goes far beyond recommendations for red-teaming practices. For example, this guidance also has developers assess potential capabilities of a future model before undergoing training or developing an incident reporting process that would be shared with third parties.

We raise this concern because the guidance that NIST is charged to develop under Subsection 4.1(a)(ii) of the EO is set to inform the reporting requirements outlined in subsection 4.2(a)(i)(C) of the EO. Subsection 4.2(a)(i)(C) requires companies that are developing dual-use foundation models to provide the federal government with reports on several of their security efforts. We believe that NIST AI 800-1 goes beyond the intent of the reporting requirements in the EO, which states that outside of reporting red-teaming testing a company should report "a description of any associated measures the company has taken to meet safety objectives, such as mitigations to improve performance on these red-team tests and strengthen overall model security." This description implies that any additional reporting would still be related to red-teaming activity, and not cover the numerous recommendations outlined in NIST AI 800-1. We urge NIST to clarify whether companies will be expected to report on all of the practices outlined in NIST AI 800-1 or only those related to conventional red-teaming testing as originally contemplated by the EO.

Responsibilities Across the AI Lifecycle

In guidance released earlier this year on generative AI models the agency noted that "not all [risk mitigation] actions apply to all AI actors. For example, actions not relevant to GAI developers may be relevant to GAI deployers. Organizations should prioritize actions based on their unique situations and context."[2] In past public comments, TechNet commended the agency's acknowledgement of actors' unique roles, and we believe it is crucial that policymakers be clear regarding the safety responsibilities of developers, deployers, and users and that these responsibilities should align with their purview through the AI life cycle.[3] In order to implement effective safety policies, it is necessary for all of these actors to have clear roles that work together in maintaining responsible AI systems.

NIST AI 800-1 should provide clearer guidelines on shared responsibilities between dual-use AI model developers and deployers. While developers can implement safeguards during model creation, they often have limited control post-deployment, especially in private, open-source, or multi-layered deployments. The current recommendations for monitoring misuse and information sharing may be impractical in many real-world scenarios. NIST should offer a comprehensive risk-

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
[2] https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf
[3] https://www.technet.org/wp-content/uploads/2024/08/TechNet-Response-to-NIST-Gen-AI-profile-1.pdf

management framework that acknowledges these limitations and suggests alternative risk mitigation strategies for developers.

Mitigation Measures for Open-Source Foundation Models

We are concerned that NIST AI 800-1's limited scope solely focusing on foundation model developers places an infeasible burden for open-source developers. The nature of open-source models limits the ability for developers to oversee downstream actors and decommission released models. In a recent op-ed, Dr. Fei-Fei Le, a renowned AI researcher from Stanford University often referred to as the "Godmother of AI," raised concerns about requiring potential "kill switches" for open-source models, stating, "If developers are concerned that [open source] programs they download and build on will be deleted, they will be much more hesitant to write code and collaborate. This kill switch will devastate the open-source community — the source of countless innovations, not just in AI, but across sectors, ranging from GPS to MRIS to the internet itself."[4] If the example safeguards outlined in AI 800-1 were to be required, it would effectively shutter American open-source development, regardless of the other risk mitigation measures a developer may take.

We advise that NIST utilize a marginal risk framework when reviewing open-source foundation models, where the benefits and risks that could come from an open-source technology are compared against those that come from other existing technologies. For example, when examining cybersecurity risks, a marginal risk analysis would consider open-source models' ability to automate vulnerability-detection tools alongside existing cybersecurity products and best practices.[5] This provides better context for the advantages and risks of new open-source models alongside the existing tradeoffs we consider for technologies. The marginal risk standard is used by other U.S. government agencies as well as by international AI Safety Institutes. The recent NTIA report on Dual-Use Foundation Models with Widely Available Model Weights adopts a marginal risk framework, stating "The consideration of marginal risk is useful to avoid targeting dual-use foundation models with widely available weights with restrictions that are unduly stricter than alternative systems that pose a similar balance of benefits and risks."[6]

Improving Misuse Identification

We believe NIST should take a more active role in curating and disseminating business-focused information on AI misuse that protects proprietary and business confidential information. This could include providing regular updates on emerging threats and detailed case studies with practical business implications so that these insights can be applied for AI developers of all sizes. NIST should also foster industry collaboration through workshops, partnerships, and sharing of best

[4] https://fortune.com/2024/08/06/godmother-of-ai-says-californias-ai-bill-will-harm-us-ecosystem-tech-politics/?abc123

[5] https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf

[6] https://www.ntia.gov/issues/artificial-intelligence/open-model-weights-report

practices, enabling businesses to collectively develop risk mitigation strategies for foundation model misuse.

NIST AI 800-1 assigns reporting responsibilities primarily to developers, but as we have noted previously, this approach may be inadequate given the complexity of AI supply chains. Foundation model developers may not always be the first to detect misuse incidents; sometimes deployers, or other entities in the supply chain will be better suited to identify and report issues.

To address these challenges, TechNet supports creating an information sharing and analysis center (ISAC). This should provide a standardized framework for defining and categorizing AI incidents, allowing for more accurate assessment and comparison of foundation model misuse across different AI systems. Such an approach would enhance NIST's ability to facilitate information sharing and collaboration across companies and actors in the AI ecosystem.

Conclusion
We look forward to working with you on AI policy and appreciate the opportunity to discuss this innovative technology. We stand ready to serve as a resource to you in your examination of this important issue. Thank you for your consideration of our perspective.

Sincerely,

Carl Holshouser
Executive Vice President