SIIA

Response to NTIA's Request for Comment Regarding "Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights" NTIA-2023-0009

Submitted by the Software & Information Industry Association

March 27, 2024

The Software & Information Industry Association (SIIA) appreciates the opportunity to provide comments on dual use foundation models with widely available model weights in response to the thoughtful request issued by the National Telecommunications and Information Administration (NTIA).[1]

SIIA is the principal trade association for companies in the business of information. Our members include roughly 375 companies reflecting the broad and diverse landscape of digital content providers and users in academic publishing, education technology, and financial information, along with creators of software and platforms used by millions worldwide, and companies specializing in data analytics and information services. Our membership includes upstream and downstream AI developers and users of AI systems in myriad environments.

SIIA's responses to the NTIA RFC reflect four core themes. First, openness should be viewed across a gradient, with model weights as one component of an AI system that can be made available to third parties in varying degrees. Second, a risk-based approach to foundation models that considers the degree and type of openness – among a broader set of considerations – would provide a stronger framework for assessing potential risks and mitigation strategies than an ex ante categorization of foundation models based on objective characteristics (such as FLOPs or whether model weights are/are not made available). Third, policymakers should take care to avoid expanding the term "dual-use foundation models" beyond the intended scope of EO 14110 based on theoretical capabilities of foundation models. Fourth, continued U.S. leadership is essential to support international alignment foundation model policy that advances innovation and aligns with core democratic values.

[1] NTIA, *Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights*, 89 Fed. Reg. 38, 14059-14063, RIN 0666-XC060, Docket No. 240216-0052 (Feb. 26, 2024) ("NTIA RFC").

1. How should NTIA define ''open'' or ''widely available'' when thinking about foundation models and model weights?

Scoping Open and Widely Available

As NTIA has rightfully noted, the terms "open" and "widely available" are "terms without clear definition or consensus" and "[t]here are gradients of 'openness,' ranging from fully 'closed' to fully 'open.'"[2] A recently proposed variant on these level-of-access gradients starts with fully closed and proceeds to hosted access, API access to a model; API access to fine-tuning; weights available; weights, data, and code available with use restrictions; weights, data, and code available without use restrictions.[3]

What this means is that considering the openness of model weights as a binary choice will not yield meaningful guidance to shape policy or best practices. Openness must be considered across a gradient or spectrum, and that requires a risk-based approach in which openness is but one factor.

In addition, openness should be considered with regard to each core artifact of an AI system. While the EO and NTIA's RFC are concerned with open or widely available model weights, openness should be considered across the AI stack, to include training data and code – and perhaps other elements that bear on transparency and the ability of end users to modify models for different applications.

Scoping "Dual-Use Foundation Model"

Many of the RFC's questions – and many of SIIA's responses below – address the qualities of open foundation models in general. The NTIA RFC and the EO are guided by a more precise term – "dual-use foundation model." A developer's obligations under section 4.2 of the EO apply only to dual-use foundation models, and the instruction to NTIA focuses on open dual-use foundation models.

The EO defines a "dual-use foundation model" in a way that is directly tied to national security. Under the EO, a covered model is one "that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security,

[2] NTIA RFC at 14061 (citing Zoe Brammer, How Does Access Impact Risk? Assessing AI Foundation Model Risk Along a Gradient of Access, The Institute for Security and Technology (December 2023); Irene Solaiman, The Gradient of Generative AI Release: Methods and Considerations, arXiv:2302.04844v1 (February 5, 2023)).

[3] Rishi Bommasani and Sayash Kapoor, et.al, Stanford University Human-Centered Artificial Intelligence, *Issue Brief: Considerations for Governing Open Foundation Models*, (Dec. 13, 2023), https://hai.stanford.edu/issue-brief-considerations-governing-open-foundation-models (hereinafter, Stanford University Human-Centered Artificial Intelligence, I*ssue Brief: Considerations for Governing Open Foundation Models*)

national public health or safety, or any combination of those matters."[4] The EO describes foundation models that are considered to be dual-use, including those that "(i) substantially lower[] the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons; (ii) enabl[e] powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or (iii) permit[] the evasion of human control or oversight through means of deception or obfuscation."[5]

The EO's examples of dual-use foundation models are not intended to be comprehensive. Indeed, the EO notes that "[m]odels meet this definition even if they are provided to end users with technical safeguards that attempt to prevent users from taking advantage of the relevant unsafe capabilities."[6] Nevertheless, this savings clause should be read in the context of the entire definition, which requires that a dual-use foundation model "exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk" to security and safety. We think a fair reading of this definition means that foundation models which are not designed or cannot easily be modified to accomplish non-civilian objectives do not give rise to serious security or safety concerns. The upshot of this is that a large language model that meets the size parameters required under the EO should not be considered dual use if designed with safeguards that would make it extremely difficult to modify to lower the barrier of entry for CBRN weapons, enable powerful offensive cyber operations, or permit the evasion of human control.

While there is value in examining the risks and benefits of large-scale foundation models that are not dual use in the way the EO defines that term, there are significant downsides to doing so in the context of implementing the dual-use foundation model requirements of the EO. Imposing the highest tier of government oversight on all large-scale foundation models would essentially subject all large foundation models to ongoing government requirements and oversight requirements under section 4.2 of the EO, and potential to export control requirements overseen by BIS. We do not think this was the intent of the EO. Indeed, doing so would limit AI innovation, research, and real-world applications without leading to measurable improvements in risk mitigation. As we describe in this submission, there are more effective ways to approach that broader universe.

1.c. Should ''wide availability'' of model weights be defined by level of distribution? If so, at what level of distribution (e.g., 10,000 entities; 1 million entities; open publication; etc.)

[4] EO 14110, Section 1(k). This is consistent but not identical with how dual use is considered under U.S. law. For example, under export control law, "[t]he term "dual-use", with respect to an item, means the item has civilian applications and military, terrorism, weapons of mass destruction, or law-enforcement-related applications." 50 USC § 4801(2).

[5] EO 14110, Section 1(k).

[6] EO 14110, Section 1(k).

should model weights be presumed to be ‘‘widely available’’? If not, how should NTIA define ‘‘wide availability?’

We recommend that NTIA propose a risk-based approach to classifying dual-use foundation models with widely available model weights. Such an approach would consider the risks, benefits, and governance of specific models that meet the definitional thresholds—both dual-use and size requirements set out in Section 4.2(b) of the EO—taking into account the degree and kind of openness. We discuss this further below in response to question 5.

2. How do the risks associated with making model weights widely available compare to the risks associated with non-public model weights?

Risks associated uniquely with open model weights stem from the potential that bad or inexperienced actors will modify weights to generate outcomes that are malicious or societally unacceptable.[7] *On the Societal Impact of Open Foundation Models*, published in February by a group of leading AI researchers,[8] explains "many of the risks described for open foundation models arise because developers relinquish exclusive control over downstream model use once model weights are released."[9]

From this starting point, the paper provides a typology of risks associated with open foundation models, including spear-phishing scams, exploitation of cybersecurity vulnerabilities, disinformation, voice-cloning scams, and the generation and dissemination of non-consensual intimate imagery (NCII) and child sexual abuse material (CSAM).[10] Crucially, the paper also concludes that some risks are more speculative than proven, such as with cybersecurity.[11] Others, such as deriving information relevant to carrying out a biosecurity attack, appear no greater than using a common internet search function.[12] Yet in some cases, such as the creation of digitally altered non-consensual intimate imagery, open foundation models do "pose

[7] See NTIA RFC at 14060 (benefits), 14061 (risks).

[8] Sayash Kapoor and Rishi Bommasani, et al., *On the Societal Impact of Open Foundation Models*, arXiv:2403.07918v1 (Feb. 27, 2024), https://arxiv.org/pdf/2403.07918.pdf (hereinafter, "Kapoor and Bommasani et al.").

[9] Kapoor and Bommasani et al. at 2. There are limitations on the ability to generalize between these two categories because openness is not an either/or option. Nevertheless, we think the "reductive" dichotomy between open and closed foundation models adopted in the recent Stanford paper provides a helpful framework to assess the risks and benefits of different foundation models. *Id.*

[10] Id. at 5-6.

[11] Id. at 7-8 (for automated vulnerability detection, "the current marginal risk of open foundation models is low and there are several approaches to defending against the marginal risk, including using AI for defense).

[12] Id. at 5.

considerable marginal risk at present."[13] Part of this is due to a lack of robust research, which in turn is partly due to the newness of these models.

Nevertheless, fully closed AI models also carry risks.[14] Closed models exacerbate certain risks because they depend on the adequacy of risk mitigation measures by the upstream developer. These models in general are less transparent, making it harder for third parties and users to detect and rectify data bias and coding anomalies.

Comparative evaluation of these risks remains an area for further research,[15] especially as the capabilities of foundation models, the uses of guardrails, and industry standards for safety and security are constantly evolving. What is clear, however, is that openness in itself, does not mean that a model can be used for malicious purposes – particularly those embedded in the EO's definition of "dual-use" – any more than a fully closed model can be used for malicious purposes.

2.b. Could open foundation models reduce equity in rights and safety impacting AI systems (e.g., healthcare, education, criminal justice, housing, online platforms, etc.)?

We believe it is a false dichotomy to posit that open foundation models – those with widely available model weights, leaving aside other assets that could be opened, and the gradient – reduce equity in rights and safety impacting AI systems vis-à-vis the "monoculture" of closed foundation models.[16] Indeed, open models provide a means to enhance equity by permitting researchers and developers to address potentially unintended biases reflected by initial model weighting *and* permitting researchers, civil society groups, and others to use powerful AI tools to investigate societal issues.

Nevertheless, we caution against a monolithic approach to open foundation models, because their potential uses, and corresponding risks and benefits, depend on a multitude of factors. Instead, as explained further in this submission, we urge NTIA to recommend a risk-based approach that includes impact on equity as an evaluation metric.[17]

[13] Id. at 7-8.

[14] See generally, IBM AI Ethics Board, "Foundation models: Opportunities, risks and mitigations," at 8-23, https://www.ibm.com/downloads/cas/E5KE5KRZ.

[15] Kapoor and Bommasani at 19-23 (Appendix containing literature review).

[16] Kapoor and Bommasani et al. at 5.

[17] We question including the term "online platform" in this question's parenthetical as on par with other areas that have received focused attention as sensitive areas for AI policy development. In addition, we encourage NTIA to provide a definition of "rights and safety impacting." While familiar with the OMB draft memo issued in November 2023, OMB has not, to our knowledge, issued a final version of that memo and we understand many commentators raised questions about the ways in which those terms were defined.

2.c. What, if any, risks related to privacy could result from the wide availability of model weights?

The wide availability of model weights is neither inherently good nor inherently bad for privacy. Foundation model developers can employ various mechanisms to protect the privacy of personal information in closed models and open models. As with our response to question 2.b, we recommend that privacy be part of any risk-based assessment of foundation models.[18]

2.e. What, if any, risks could result from differences in access to widely available models across different jurisdictions?

We believe there is risk in adopting fundamentally different approaches to oversight of foundation models of any type across different jurisdictions. Jurisdictions will inevitably have local approaches to different issues that warrant specific attention, but in terms of broad guidance, innovation, and responsible development would benefit from consistency. As described below, we urge the U.S. government to continue to exercise leadership in the global community.

3. What are the benefits of foundation models with model weights that are widely available as compared to fully closed models?

On the benefits side of the equation, openness has already proved to be a catalyst for research and innovation by essentially democratizing access to models that are cost-prohibitive for many actors in the AI ecosystem to develop on their own.[19] Open access allows for customizability, which permits downstream developers to diversify model behavior. Closed models may also allow customizability through fine tuning APIs, for example, but may limit the degree of customizability. This is among the risk/benefit considerations that are nearly impossible to determine in the abstract and require an approach that considers not only the degree of openness but other specifics of each model. The same can be said for benefits around safety and security and the degree to which models permit researchers and downstream developers to interrogate models and address vulnerabilities during development or deployment. For these reasons, as we describe below, we endorse a risk-based approach.

In addition, we support efforts of the U.S. government to advance further ways to democratize AI research and innovation through the NAIRR pilot, the AI Institutes program led by NSF, and other initiatives. We also support the efforts of private industry to make models that meet standards for responsible development (including red-teaming, capability evaluation, TEVV measures, and so on).

[18] See, e.g., Bandan Chandra Das, et al., *Security and Privacy Challenges of Large Language Models: A Survey*, arXiv:2402.00888v1 (Feb. 2024), https://arxiv.org/pdf/2402.00888.pdf.

[19] This is summarized in the NTIA RFC's discussion of benefits and risks. NTIA RFC at 14060 (benefits), 14061 (risks).

3.a. What benefits do open model weights offer for competition and innovation, both in the AI marketplace and in other areas of the economy? In what ways can open dual-use foundation models enable or enhance scientific research, as well as education/ training in computer science and related fields?

Open model weights and other mechanisms for making accessible highly capable AI tools help to reduce market concentration in downstream development. This has enormous positive effects on society. As noted in the Kapoor and Bommasani paper, openness "may yield more diverse down-stream model behavior," reducing certain types of risks..[20] Researchers and downstream developers are able to use open models to develop new applications and investigate critical questions without having to front the enormous capital costs associated with foundation model development.

3.c. Could open model weights, and in particular the ability to retrain models, help advance equity in rights and safety-impacting AI systems (e.g., healthcare, education, criminal justice, housing, online platforms etc.)?

Please see above our response to question 2.b on the risks to equity.

3.d. How can the diffusion of AI models with widely available weights support the United States' national security interests? How could it interfere with, or further the enjoyment and protection of human rights within and outside of the United States?

National security implications of open models are potentially immense although, again, it depends on many factors. The potential benefits are to create models that are highly adaptable for different national security missions, an area that based on public reporting is already well underway at the Pentagon and beyond. In addition, openness of models from U.S. developers has a "soft power" benefit by further innovation beyond U.S. borders and providing the diffusion of models that are not pre-set with potential vulnerabilities or biases that may be required of models released from developers in certain other countries. As to the impact on human rights, we believe this is similar to the impact on equity – it necessarily depends; there are clear potential benefits to openness and transparency, but also risks that must be managed.

4. Are there other relevant components of open foundation models that, if simultaneously widely available, would change the risks or benefits presented by widely available model weights? If so, please list them and explain their impact.

Model weights are but one artifact of an AI system that a developer can open to broader usage. Other artifacts include training data and code. For each artifact, openness can occur across a

[20] Kapoor and Bommasani et al. at 5.

gradient. For example, licensing terms and use restrictions impact the degree of "openness" with corresponding implications for potential risks.[21]

5. What are the safety-related or broader technical issues involved in managing risks and amplifying benefits of dual-use foundation models with widely available model weights?

Risk management is at the heart of responsible AI, no less in open models than in closed models. Those risks can be mitigated through various measures, including staged release; less than fully open access; limitations on who can access the weights (e.g., through licenses and user restrictions); and limitations on how the assets can be used (e.g., use restrictions and contract terms).[22] Moreover, open innovation in the AI context creates opportunities to mitigate risk through the efforts of researchers and downstream developers to interrogate and improve open models.

What is clear, however, is that the principles that ground responsible AI must apply as well to generative AI models, including those that offer fully or partially open assets such as model weights. Industry and experts are virtually unified in endorsing a risk-based approach to AI development and deployment. We encourage NTIA to endorse such an approach in the context of foundation model governance (across the gradient of openness) and, in particular, for dual-use foundation models.

NIST provides the starting point for this endeavor. Building on the seminal AI Risk Management Framework released in January 2023,[23] the NIST AI Safety Institute is now focused on an effort to develop a companion resource for generative AI.[24] SIIA is engaged in this work as a member of the AI Safety Institute Consortium. The work of NIST has been critical in raising the bar on responsible AI and moving the industry in a positive direction. We see evidence of this in the

[21] See Stanford University Human-Centered Artificial Intelligence, *Issue Brief: Considerations for Governing Open foundation Models* (Dec. 2023), at 3-4. https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf.

[22] See, e.g., https://ai.google.dev/gemma/prohibited_use_policy. Google's release of Gemma – an open model that falls under the size threshold of section 4.2 of EO 14110 – illustrates a responsible, transparent approach to public release, reflect in the accompanying model card (https://www.kaggle.com/models/google/gemma) and technical report (https://storage.googleapis.com/deepmind-media/gemma/gemma-report.pdf).

[23] NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (January 2023). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[24] NIST, *US Artificial Intelligence Safety Institute: AISIC Working Groups*, (Nov. 17, 2023). https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute/aisic-working-groups.

growing efforts taken by upstream developers of open models to mitigate risks before making models available.[25]

Nevertheless, in the context of open models, it is important to recognize that responsibility for mitigating risks must be undertaken not only by upstream developers but also by downstream developers and sophisticated deployers. Once artifacts of a model are made open, it is not possible for an upstream developer to claw them back. That means it is important to have principles of responsible use and guardrails to prevent improper use.

The Kapoor and Bammasani paper provides a framework for risk assessment in the context of open models that we believe is instructive and should guide recommendations in this space. The framework would require evaluation of potential threats, risks and defenses to those risks in the absence of open foundation models, evidence of marginal risk of open foundation models versus closed models, ease of defending against new risks identified, and limitations in any analysis.[26] Assessments along these lines – by both upstream and downstream developers – are in line with best practices and supplement other risk mitigation strategies employed for AI systems in general.[27] In addition to assessments, as researchers at Berkeley recently explained, there are steps that model developers can undertake, such as staging the release of different model artifacts, that can help to insulate against some of the risks inherent in foundation models.[28]

In short, we caution against a one-size-fits-all approach to mitigating risks for open models due to the gradient of openness, the differences among models, and differences around model training data. In addition, advances in foundation models, risk mitigation techniques (TEVV, auditing, red-teaming, and so on) and the capabilities of bad actors mean that any approach must be sufficiently flexible and agile to adapt.

Nevertheless, there are discrete areas in which the risk associated with foundation models and in particular open models may be sufficiently acute that action by Congress would be necessary. Among these we would urge Congress to criminalize the creation and dissemination of NCII to raise the cost on bad actors. This is an area that the researchers in the Kapoor and Bommasani

[25] Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House.

[26] Kapoor and Bommasani et al. at 5-7.

[27] We also recommend the work of Stanford's HAI, the Partnership on AI, and Berkley's GPAIS. See https://hai.stanford.edu/sites/default/files/2023-12/Governing-Open-Foundation-Models.pdf;

https://partnershiponai.org/wp-content/uploads/1923/10/PAI-Model-Deployment-Guidance.pdf;

[28] UC Berkeley Center For Long-Term Cybersecurity, *AI Risk-Management Standards Profile for General-Purpose AI Systems (GPAIS) and Foundation Models*, (Nov. 2023). https://cltc.berkeley.edu/wp-content/uploads/2023/11/Berkeley-GPAIS-Foundation-Model-Risk-Management-Standards-Profile-v1.0.pdf.

paper identify as having greater marginal risk in open versus closed models, and an area that is not covered well by existing law.[29]

6. What are the legal or business issues or effects related to open foundation models?

6.a. In which ways is open-source software policy analogous (or not) to the availability of model weights? Are there lessons we can learn from the history and ecosystem of open-source software, open data, and other ''open'' initiatives for open foundation models, particularly the availability of model weights?

The overall lesson of the open source software movement has been that openness has led to greater rather than less safety, as well as innovation, transparency, and collaboration.[30] This is well documented and indeed recognized by U.S. government policy.[31] We believe the future of AI will trend towards openness, though the risks of an AI model versus software are different.

6.b. How, if at all, does the wide availability of model weights change the competition dynamics in the broader economy, specifically looking at industries such as but not limited to healthcare, marketing, and education?

The availability of a broad range of foundation models from companies, both large and small, has created a fiercely competitive market. Many companies have been working on foundation models for years, others are new entrants, some models are open-source, others restricted, and there is no way to know which model(s), if any, ultimately will be the most successful. One thing seems certain, though, big is not necessarily better. This is because no foundation model is able to do everything well. Rather, different models work better for different tasks and use cases. A model developed to help doctors create better and more personalized patient treatment plans, for example, will not be much help to a finance professional looking to draw insights from reams of financial data, or a manufacturing company trying to optimize its production processes. They will need different models customized to their particular needs.

Barriers related to the resources, time, and cost, it takes to build, train, and deploy models have also been significantly reduced. Customers have many choices for sourcing their compute capacity needs. The cloud offers great flexibility and scalability, but there are also various on-

[29] Christopher A. Mohr, President, Software & Information Industry Association, *Artificial Intelligence and Intellectual Property, Part II – Identity in the Age of AI*, Statement Before the House Committee on the Judiciary, Subcommittee on Courts, Intellectual Property and the Internet, (Feb. 2, 2024). https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/mohr-testimony-1.pdf

[30] Guido Schryen, Is Open Source Security a Myth?, *Communications of the ACM*, 54(5), (May 2011). https://dl.acm.org/doi/pdf/10.1145/1941487.1941516

[31] See, e.g, The Digital Services Playbook Play 13: https://playbook.cio.gov/#play13 and The Department of Defense Open Source Software FAQs: https://dodcio.defense.gov/Open-Source-Software-FAQ/.

premises solutions, as well as hybrid solutions combining these and other options. Increasing access to quality data, and the availability of pre-trained models that are customizable, also help lower barriers to entry, which makes it easier for new entrants to compete with more well-established companies.

All told, these dynamics encourage innovation and increase competition to the benefit of the broader economy.

6.d. Are there concerns about potential barriers to interoperability stemming from different incompatible ''open'' licenses, e.g., licenses with conflicting requirements, applied to AI components? Would standardizing license terms specifically for foundation model weights be beneficial? Are there particular examples in existence that could be useful?

Historically speaking, developers of technology do so to ensure their products in their developer suite operate well but it has often been difficult for interoperability between different developers. There are business and other reasons to limit interoperability, but interoperability can be beneficial for the broader AI ecosystem. Industry best practices, standardized licensing terms, and other efforts can help to promote interoperability across models from different upstream developers.

7. What are current or potential voluntary, domestic regulatory, and international mechanisms to manage the risks and maximize the benefits of foundation models with widely available weights? What kind of entities should take a leadership role across which features of governance?

In general, we believe that alignment across jurisdictions is essential for innovation; that guardrails around AI designed to mitigate the potential downsides should, to the extent possible, be consistent; and that government work towards agile, flexible approaches that can adapt as the risks and benefits of technology inevitably develop (indeed, ChatGPT has been part of public discourse for only 18 months, and the European Commission's proposal for the AI Act included no provisions on "general purpose AI").

There have been significant undertakings to mitigate risks and maximize the benefits of foundation models of all types, and we encourage NTIA to emphasize the value of these efforts for creating a safe and secure ecosystem and one that maximizes international alignment.

Among others, we highlight the Voluntary Commitments agreed to by the White House and leading AI developers;[32] the G7 International Code of Conduct[33] and continuing work of the G7;

[32] The White House, *Voluntary AI Commitments*, (September 2023). https://www.whitehouse.gov/wp-content/uploads/2023/09/Voluntary-AI-Commitments-September-2023.pdf

[33] G7 2023 Hiroshima Process, *International Code of Conduct for Organizations Developing Advanced AI Systems,* (Oct. 30, 2023). https://www.soumu.go.jp/main_content/000912748.pdf

industry-led efforts at the Partnership on Artificial Intelligence[34] and the AI Verify Initiative;[35] and ongoing work by the OECD and the Global Partnership on AI.[36] These efforts are aligned in seeking to raise the bar on developers across the AI value chain to implement best practices for safe, secure, and trustworthy foundation models at the same time as the technology is advancing at a rapid pace.

Specifically with regard to open models, however, as the Kapoor and Bommasani researchers point out, some of these efforts focus specifically on closed foundation models and if they "are interpreted strictly to apply to foundation model developers, independent of how the model is adapted or used downstream, they would be difficult for open developers to comply with."[37] Therefore, as the U.S. government continues efforts to negotiate the ground rules for foundation models, it should also give attention to the unique relationship between open models and downstream developers and apportion responsibilities between these parties appropriately.

Overall, we encourage the U.S. government to exercise even more leadership in the international community. The United States and China are the two leading homes of AI researchers and companies in the world and we believe the United States' imprint on international AI policy is essential to ensure that the technology advances in a manner that is aligned with democratic values.

7.b. How might the wide availability of open foundation model weights facilitate, or else frustrate, government action in AI regulation?

We encourage the government to focus efforts on use-based restrictions to target those activities that are deemed harmful. On the development side, continuing attention to best practices, industry standards, and robust risk-management assessment processes should remain the focus. Direct government oversight of foundation models should be limited to those that meet a narrow definition of dual-use as described earlier in this submission.

7.d. What role, if any, should the U.S. government take in setting metrics for risk, creating standards for best practices, and/or supporting or restricting the availability of foundation model weights?

We endorse the ongoing work of the NIST AI Safety Institute and Consortium to develop a companion to the NIST AI Risk Management Framework. We anticipate this will address the very

[34] Partnership on AI, *PAI's Guidance for Safe Foundation Model Deployment*, (October 2023). https://partnershiponai.org/wp-content/uploads/1923/10/PAI-Model-Deployment-Guidance.pdf

[35] AI Verify Foundation, *Proposed Model AI Governance Framework for Generative AI: Fostering a Trusted Ecosystem,* (Jan. 16, 2024). https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf

[36] OECD Legal Instruments.

[37] Kapoor and Bommasani at 9.

question presented here and would encourage the U.S. government to avoid conflicting guidance.

7.i. Should other government or nongovernment bodies, currently existing or not, support the government in this role? Should this vary by sector?

Please see above our answer to question 7.

7.g. What should the U.S. prioritize in working with other countries on this topic, and which countries are most important to work with?

The U.S. should prioritize innovation and promotion of American industry as leaders in ensuring transparency and best practices with regard to the development of responsible AI systems and foundation models to benefit society at large. We encourage the U.S. government to advocate for cross-collaboration among AI safety institutes globally to develop standards as well as collaborate with industry in these discussions to develop appropriate and effective policies.

8. In the face of continually changing technology, and given unforeseen risks and benefits, how can governments, companies, and individuals make decisions or plans today about open foundation models that will be useful in the future?

8.a. How should these potentially competing interests of innovation, competition, and security be addressed or balanced?

If we have learned anything from the past, technology is here to stay. We need to learn how to adapt our systems to meet the present and future of technological advancement and innovation. Technical policy expertise is needed on the federal level in order to adapt to emerging needs. Policies should be amenable to different use cases and creating incentives to allow companies to innovate and maintain security are essential to create a collaborative ecosystem between industry and government.

8.b. Noting that E.O. 14110 grants the Secretary of Commerce the capacity to adapt the threshold, is the amount of computational resources required to build a model, such as the cutoff of 10^{26} integer or floating-point operations used in the Executive order, a useful metric for thresholds to mitigate risk in the long-term, particularly for risks associated with wide availability of model weights?

Thresholds should be based on qualitative metrics rather than FLOPs. Technology will continue to advance, and it is possible that models in excess of the threshold will present less risk than models that fall below the threshold. We encourage the Secretary of Commerce to work closely with the NIST AI Safety Institute to develop appropriate guidance for assessing risk and imposing oversight obligations under the EO.