ITI Promoting Innovation Worldwide

Mr. Travis Hall
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Ave NW
Washington, DC 20023

March 27, 2024

Re: ITI Feedback to NTIA RFC on Dual Use Foundation AI Models with Widely Available Model Weights (Docket No. 240216-0052; RIN# 0660-XC06)

Dear Mr. Hall,

The Information Technology Industry Council (ITI) welcomes the opportunity to provide feedback to the National Telecommunications and Information Administration (NTIA) **Request for Comment (RFC) Related to NTIA's assignment Under Sections 4.6 (a) of the Executive Order Concerning Artificial Intelligence on dual use foundation artificial intelligence models with widely available model weights (E.O.).**

ITI represents the world's leading information and communications technology (ICT) companies. We promote innovation worldwide, serving as the ICT industry's premier advocate and thought leader in the United States and around the globe. ITI's membership comprises leading innovative companies from all corners of the technology sector, including hardware, software, digital services, semiconductor, network equipment, and other internet and technology-enabled companies that rely on ICT to evolve their businesses. Artificial Intelligence is a priority technology area for our member companies, who are both developing and using the technology to evolve their businesses.

ITI is committed to fostering the responsible development and deployment of AI. We have been actively engaged in shaping AI policy around the world. In 2021, we issued a set of *Global AI Policy Recommendations*, aimed at helping governments facilitate an environment that supports AI while simultaneously recognizing that there are challenges that need to be addressed as the uptake of AI grows around the world.[1] We also launched our AI Futures Initiative in 2023, an initiative comprised of technical and policy experts aimed at addressing challenging questions that are emerging in the global conversation on AI. We have published several policy papers via this Initiative, including on the *AI Value Chain and Foundation Models*, and on *AI-Generated Content Authentication*, both of which we think are particularly relevant to NTIA's RFI.[2] We have also actively worked to inform the efforts of the National

[1] Our complete *Global AI Policy Recommendations* are available here: https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf

[2] ITI's *Guide to AI Content Authentication* available here: https://www.itic.org/policy/ITI_AIContentAuthorizationPolicy_122123.pdf and ITI's *Understanding Foundation Models & the AI Value Chain* paper available here: https://www.itic.org/documents/artificial-intelligence/ITI_AIPolicyPrinciples_080323.pdf

Institute of Standards and Technology (NIST)[3] to create an AI Risk Management Framework (RMF) and have consistently contributed to the debate in the EU on its AI Act.

General Feedback

Below, we offer several high-level points, followed by answers to more specific questions that NTIA poses in the RFC.

- **NTIA should consider closed and open models to include a spectrum of foundation models**. We appreciate how NTIA highlights the difference between open and closed foundation models by citing relevant examples for each category. NTIA should also note that some foundation models do not neatly fall into either category, and can include other models that may have features of either broad group. As we explain below in our response, experts classify openness on a gradient with the following categories: Fully closed, hosted access, API access to model, API access to fine tuning, weights available, weights, data, and code available with and without use restrictions.[4]

  Furthermore, the choice between leveraging open or closed models presents its own set of considerations. Open models or foundation models with widely available model weights foster collaboration and rapid iteration because they are accessible and customizable by a wide community of developers and researchers. On the other hand, closed models, which are proprietary and tightly controlled, might be preferred in scenarios where data privacy or intellectual property concerns are paramount.

- **NTIA should consider that access to different component parts of open foundation models may change the risk and benefit calculus.** For example, access to model weights alone may present a limited risk, while access to model weights plus source code could marginally increase risks and benefits (such as enhancing research and transparency) because it could grant a user the ability to make more significant changes to the model. For more sophisticated actors, the ability to use other existing source code to finetune a model will remain regardless of whether the code is shipped with the weights.

- **NTIA should clarify that not all open foundation models or foundation models with widely available weights are dual use foundation models as defined by the E.O**. In considering open foundation models / foundation models with widely available weights, NTIA should recognize that not all of them possess the characteristics or capabilities outlined in the E.O., which constitute a dual-use foundation model. Treating all open foundation models as dual use foundation models with widely available weights would adversely impact innovation and competitiveness of many actors across the AI ecosystem.

[3] See ITI response to RFI on AI RMF Concept Paper here: ITI Comments on AI RMF Concept Paper FINAL.pdf

[4] Bommasani, Rishi, et al. "Issue Brief Considerations for Governing Open Foundation Models | Stanford HAI." Stanford University Human-Centered Artificial Intelligence, 13 Dec. 2023, https://hai.stanford.edu/issue-brief-considerations- governing-open-foundation-models.

- **NTIA should adopt a risk-based approach with respect to open foundation models since not all models pose an equivalent degree of risk.** In considering policy options to present in its report, NTIA should at this stage focus on open foundation models that rise to the level of dual use – these are foundation models that may pose a serious security risk, a risk to national/economic security, or a risk to national public health and safety. A 'one size fits all' approach would disproportionately impact researchers, academics and developers of open foundation models that do not rise to the level of dual use. Broader policy options for open foundation models should be considered once an appropriate approach to closed models, and any marginal risk posed by open models, are better understood.

- **NTIA should acknowledge that risk management is a shared responsibility across the AI value chain.** This approach is consistent with the way we have discussed the AI value chain in our *Understanding Foundation Models & the AI Value Chain* paper, which reflects the need for different actors to share responsibility across the value chain. This is especially important because once foundation models with widely available weights are deployed, developers are not able to retract said model, even in cases where that model is being used in malicious ways.

- **NTIA should work with stakeholders across the AI value chain.** NTIA would benefit from hearing from a range of diverse stakeholders in the AI value chain, including developers, deployers and end users. Foundation models with widely available weights and closed models can be used and deployed under various contexts. There are clearly downstream effects when deploying these models, which is why it is important for NTIA to consult with all stakeholders across the AI value chain.

- **NTIA should prioritize evidence based and scientifically informed policymaking while considering recommendations for dual use foundation AI models with widely available weights.** Any recommendations addressing risks of dual use foundation models with widely available weights should follow guidance developed by NIST through the U.S. AI Safety Institute, which is tasked with studying and identifying mitigations to AI safety risks under the E.O.

- **NTIA, and the USG more broadly, should ensure robust international cooperation and coordination**. We appreciate that the overarching E.O. from which the RFC stems recognizes the importance of working with key international allies to further bolster innovation and address risks. The U.S. should remain engaged in AI policy discussions in both bilateral and multilateral fora, and look to progress discussions around open foundation models because this is a conversation that is taking place across jurisdictions.

Specific Responses

Please find below our response to several of the questions posed by NTIA.

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

In defining "open" foundation models, we encourage NTIA to consider that there is a gradient of "openness" for foundation models. Experts classify this gradient using the following categories: fully closed, hosted access, API access to model, API access to fine tuning, weights available, weights, data and code available with and without use restrictions.[5]

We appreciate that later in the RFC, NTIA specifically asks a question about if and how risks vary depending on which components of an AI model are available (e.g. model weights, model weights AND source code, model weights, source code, AND data sets, etc), which demonstrates to us that there is in fact a recognition that the risk a model poses depends on the interplay of the three elements. That said, we think it is worth reiterating again and encourage NTIA to clearly explain this in its forthcoming report.

> *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.) should model weights be presumed to be "widely available"? If not, how should NTIA define "wide availability?"*

In thinking through how to appropriately define "widely available," we encourage NTIA to consider whether "widely available" is actually a realistic proxy for risk. "Wide availability" as a metric may inadvertently overlook the functionality of the model, which may also present risk in a way that simple availability does not. To be sure, a model could be widely available but present limited risk; or, it could have limited availability but be significantly risky. The definition should ideally reflect a balance that allows for broad use, competition, and innovation while managing risks associated with the deployment of the specific AI model.

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

ITI and its members acknowledge that risks arise from the malicious use of both widely available and non-public model weights. We explore several of these risks in our paper on *Understanding Foundation Models and the AI Value Chain.*[6] It is important to note that in many instances, risks arising from the introduction of models with open weights and risks arising from models with closed weights are not significantly different. Risks arising from malicious use exist; however, more research and analysis is required to ground policy interventions in a way that will help address said risks.

The risk calculus may change because of the ease with which a malicious actor may be able to access and fine-tune an open model, where it might be more difficult to leverage a closed model for malicious purposes given said models cannot be fine-tuned or adapted as easily (though this not to say that they *cannot* be). Additionally, once an open model is released, the developer no longer has control over how it is used – it cannot be pulled back.

[5] Bommasani, Rishi, et al. "Issue Brief Considerations for Governing Open Foundation Models | Stanford HAI." Stanford University Human-Centered Artificial Intelligence, 13 Dec. 2023, https://hai.stanford.edu/issue-brief-considerations- governing-open-foundation-models.

[6] ITI's *Understanding Foundation Models & the AI Value Chain* paper available here: https://www.itic.org/documents/artificial-intelligence/ITI_AIPolicyPrinciples_080323.pdf

Although open models may, in certain instances, present an increased risk due to their unique characteristics, it is important to determine if and how that risk differs from existing risks (and if and how it differs from risks presented by closed models). The Framework introduced by Stanford for measuring the marginal risk of open foundation models is something that may be worth further exploring.[7] Overall, any policy interventions and governance models should be dynamic and adaptable to the evolving landscape of AI technology.

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

    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?*

There is growing evidence[8] about the benefits of open foundation models. This includes enabling competition, catalyzing innovation and facilitating transparency.[9] Additionally, widely available model weights can help to democratize access and use of AI systems, allowing a greater number of users to contribute to AI development processes.

Widely available model weights also allow for independent evaluations of software by a wider community of developers, enabling them to identify vulnerabilities and test for safety issues. This openness promotes the adoption of robust cybersecurity practices since a diverse pool of experts are able to effectively evaluate model components (training data, model weights, source code) to mitigate risks that may otherwise go unnoticed, which in turn can lead to improved and safer foundation models. These practices can also support the establishment of verifiable benchmarks for model performance in safety and compliance.

Widely available model weights may also reduce the environmental impact of training large models, which use significant amounts of energy. Open model weights can also promote a more competitive and diverse ecosystem by reducing market concentration and barriers to entry.

Further, fundamental research in AI relies on transparent collaboration in the development of models as a primary method for research communities to collaborate. Models with open weights, code and data have become as important to AI research as traditional publication and conferences because they provide reference implementations and examples of research, and testbeds for collaboration in the research community. For example, the field of large language models (LLMs) began with open,

[7] Kapoor and Bommasani et al, On the Societal Impact of Open Foundation Models, Stanford HAI, 2023.

[8] Seger, Dreksler, Moulange et al, Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives, Centre for the Governance of AI, 2023. Available here: https://cdn.governance.ai/Open-Sourcing_Highly_Capable_Foundation_Models_2023_GovAI.pdf.

[9] Kapoor and Bommasani et al, Considerations for Governing Open Foundation Models, Stanford HAI, 2023. Available here: https://hai.stanford.edu/issue-brief-considerations-governing-open-foundation-models

collaborative research including the seminal BERT model that was an implementation of the new Transformer approach that is considered the critical breakthrough that made LLMs possible. BERT was released under the MIT open-source license.

Finally, widely available model weights are able to promote innovation allowing actors across the AI value chain to create new economic opportunities in diverse fields such as marketing, communications, medicine, education, and employee training. Developers and deployers can customize their models depending on the specific use case.

b. How can making model weights widely available improve the safety, security, and trustworthiness of AI and the robustness of public preparedness against potential AI risks?

Foundation models with widely available model weights provide upstream and downstream users with the ability to make more informed decisions and effectively work with the models that they are training. Additionally, with access to weights and other components, developers and deployers can adapt models more quickly, therefore saving time and resources. This makes using foundation models more cost-effective, business friendly, and allows for a broad range of applications.

Widely available model weights can also help facilitate transparency. In line with ITI's prior positioning (see, for example, our paper on the *AI Value Chain & Foundation Models*[10]), transparency is important throughout the AI value chain, and is especially important to risk management. Access to model weights can help to provide insight into the ways in which certain models were developed. Researchers can use the weights to analyze the strength of connections between different components or links within the model, thus potentially revealing information about how the model prioritizes various factors.[11]

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.)?

Weights can help to influence the performance of AI model outcomes. As weights are fine-tuned, so are the accuracy of outcomes. Having weights available makes retraining models more cost effective for users and can promote more accurate models. Additionally, testing also plays a significant role in improving model performance and accuracy.

We also stress that more expertise is required to work with model weights alone, which may serve to limit the circle of capable developers. Along the gradient of openness, a model with only adjustable weights is limited in customizability compared to a model which is radically open, meaning that anyone has full access to all components of including data, weights, and source code.

[10] See ITI's *Understanding Foundation Models & The AI Value Chain* paper here: https://www.itic.org/documents/artificial-intelligence/ITI_AIPolicyPrinciples_080323.pdf

[11] Prompt Engineering Institute. (2023, December 1). Openness in Language Models: Open-Source vs Open Weights vs Restricted Weights. https://promptengineering.org/llm-open-source-vs-open-weights-vs-restricted-weights/

In areas where concerns surrounding sensitive data processing and security are most acute, NTIA should consider how models with widely available weights could be beneficial. As discussed above, these types of models are oftentimes more conducive to operating within a given organization or business' network, which would allow the model to operate in a more isolated environment. This would make it easier to isolate the model's processing of sensitive data (personal or other proprietary data) to within a company's own systems and to monitor that environment for any communication attempts made to external systems, which is indicative of command and control or data exfiltration behaviors.

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.**

The risk-benefit landscape can be altered by considering other relevant components including training data, source code and model architecture, evaluation metrics, usage guidelines and pre-trained data.

a. Training data: Making datasets that AI models are trained on (open) can increase transparency. However, at the same time, it is important for companies to manage this data carefully to protect personally identifiable information and prevent misuse.

b. Model architecture and source code: The availability of these components can accelerate innovation by enabling developers and deployers to customize and improve upon existing models.

c. Evaluation metrics: Broadening access to evaluation criteria can lead to a more standardized and transparent methodology in assessing AI performance.

d. Documentation and usage guidelines: Their availability can guide responsible use, helping users understand the models' capabilities and limitations, and prevent misuse.

e. Pre-trained AI models: The availability of pre-trained models, beyond just the weights, includes configurations and tuning parameters. This can significantly lower the barrier to entry for using advanced AI technologies, democratising access.

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?**

What model evaluations, if any, can help determine the risks or benefits associated with making weights of a foundation model widely available?

Best practices for ensuring safety of AI models may include red teaming AI models, or testing models for flaws or other vulnerabilities before they are released. The recently launched U.S. AI Safety Institute will be an important venue to facilitate collaboration on ways to test, evaluate, verify, and validate open models, including establishing guidelines to help determine the benefits and risks associated with widely available model weights.

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

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 history of open-source software and other "open" initiatives can, in certain ways, provide helpful reference points for the conversation taking place around open model weights. There are a few key themes that underpin the open-source software conversation that may provide useful analogs or at least a conceptual frame for open model weights.

- **Transparent Collaboration.** The concept of transparent collaboration underpins open-source software. Software developers with a variety of affiliations, such as commercial, academic, and government, as well as individuals, can together access, modify, and contribute source code to open-source codebases which can drive rapid innovation, be critical in catching and fixing bugs, and improve the overall functionality of software. In the same way that transparent collaboration is helpful to improving functionality in the open-source context, it can also be a useful concept for the development of foundation and other machine learning models. While source code may not always be directly available, allowing a wide variety of users access to at least the weights of a model can help in accelerating innovation, spotting issues, patching vulnerabilities, and improving functionality of the model.

- **Research**. A de-facto method for research communities in artificial intelligence and other areas to collaborate world-wide, and especially in the United States, has been via open-source software - to provide examples, working implementations and testbeds for collaborative development of research, and as such open source has become as important to research as traditional publication and conferences. Research in artificial intelligence also relies on transparent collaboration in the development of models. As such, it is critical that any regulation relating specifically to models with widely available weights not slow down or stifle the flow and exchange of information vital for research.

- **Cybersecurity.** The security of open-source software is an evolving conversation. A few months ago, the Office of the National Cyber Director requested information on open-source software security so that it can further implement Section 4.2.1 of the National Cybersecurity Strategy, aimed at improving open-source software security. In particular, ONCD recognizes the immense benefits of open-source software, but also highlights the possible negative downstream impact of vulnerabilities. Just as in the case of open source software, the openness of a model does not make it inherently more or less secure than a closed access model, and so while openness may mean that the model is more vulnerable to exploitation and modification for malicious purposes, especially if source code and training data is available, it also means that these vulnerabilities can be found and fixed proactively, and exploits rapidly addressed, by the developer community, analogous to how cybersecurity issues are often handled by open source communities because of the ability to transparently collaborate on addressing the problem.

- **Licensing.** The open-source software community relies heavily upon a common licensing framework to grant developers the right to access, modify, and distribute the code. While the specific licensing regime used in the open-source software context is not directly portable to an open model weights context, the overall concept of a licensing regime remains applicable. For open model weights, it is worth considering what a framework would look like and how terms of use will be discovered. We discuss this further below.

b. 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?

Standardizing license terms for foundation models could be an important component of a holistic approach to open model safety. Licenses can preserve the benefits of open models and help support responsible use. While licenses can help to dictate the terms and conditions of use, as well as the ability of a user to redistribute the software, they are not a silver bullet solution. To be sure, licenses will not prevent a malicious actor from leveraging the model for nefarious purposes.

That being said, a licensing framework may be helpful in providing guidance and fostering a common understanding of terms and conditions for developers looking to open their models. Importantly, the ongoing conversation about if and how traditional software licenses might apply to open model weights is ongoing. To be sure, open model weights are in many instances released under existing licensing mechanisms, like MIT or Apache 2.0.[12] However, we think it important to point out that traditional software licenses may not be directly applicable when adapted for AI models. For example, it may be challenging to apply traditional open-source software licenses to AI models because they do not entirely account for the technical nature and capabilities of an AI model. Therefore, an effective licensing framework may need to encapsulate more components of a model to include things like weights, but also data, and other factors. We do not believe that NTIA should be charged with designing a licensing regime, but offer the below thoughts which may shape a future community framework for AI models with widely available weights:

- Training & Deployment

Deployment code and training code are the building blocks of AI models. A license for open models should consider the use, modification, and distribution of this code.

- Data & Weights

Databases can be covered by a variety of licenses depending on the copyright and public availability of the data. Specific database licenses are also a possibility. Weights are not as simple to license as some models provide access to weights where others do not. Licenses for fully open models must consider the data and weights with which a model is trained as they are important for replicating the model's performance.[13]

[12] Hugging Face's model index provides an overview of various licenses that model weights are released under, available here: https://huggingface.co/models?sort=trending

[13] Riddiugh, J. (2024, March 7). Licensing and legal considerations for open-source AI.

- Impact, Use, & Documentation

Licenses can be designed to take into account the impact, use, and documentation of a model. License designs should consider potential unwanted side effects, however, in particular, that use restrictions being proposed for open access model licenses do not hinder the ability for commercial enterprises to use and contribute to such models.

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 weight? What kind of entities should take a leadership role across the features of governance?**

NTIA may want to consider highlighting in its report the significant role that stakeholders can play in establishing a community framework to contribute to the responsible development of open models.'[14]

- *Platform for sharing best adoptions of models with widely available weight*

A community-based platform for gathering feedback from users and developers would assist in promoting the benefits of open sourcing AI and lead to more optimal model development.

- *Clear guidelines addressing ethical use*

Ethical use guidelines are an essential part of an open model community framework. Such guidelines must address the responsible use of open models and include principles on issues such as bias and privacy.

- *Comprehensive licensing information on all components of a model*

Open-source AI licenses should consider all components of AI models including the code, training data, model weights, and use of a model.

NTIA should also prioritize engagement with stakeholders to discuss approaches to and governance issues related to open foundation models. This conversation is taking place across jurisdictions, and it would behoove NTIA, and the U.S. government as a whole, to leverage existing multilateral dialogues such as the G7 process, forums held around the forthcoming AI Safety Summits. and the UN AI Advisory Board.

b. When, if ever, should entities deploying AI disclose to users or the general public that they are using open foundation models either with or without widely available weights?

As we highlight in ITI's *Policy Principles for Enabling Transparency of AI Systems*[15], disclosure generally refers to making a user aware of the fact that they are interacting with or using an AI system – usually in real time or during the use of the system. We encourage policymakers to take a risk-based approach to

[14] Yadav, R. (2023, December 18). #117 flexing open weights. Exploring the Evolution of Open-Source Models in the AI Era: From Open-API to Open Weights.

[15] See ITI's *AI Transparency Policy Principles* available here: https://www.itic.org/documents/artificial-intelligence/ITIsPolicyPrinciplesforEnablingTransparencyofAISystems2022.pdf.

disclosure requirements, focusing on disclosure in high-risk settings. These principles should apply irrespective of whether they are interacting with open or closed foundation models.

> *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?*
>
> > *i. Should other government or non-government bodies, currently existing or not, support the government in this role? Should this vary by sector?*

NIST and international standards organizations can play a role in supporting international standards that are aimed at setting metrics for risk management and developing best practices and guidelines for the development and release of foundation models.

E. 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 foundation models that will be useful in the future?

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 1026 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?

While we appreciate that the Commerce Department (and the Administration more broadly) needed to start from somewhere in defining what constitutes a dual-use foundation model, we have concerns about using floating point operations (FLOPs) as a useful way to classify risk. Compute is not necessarily indicative of risk, in the same way that wide availability does not necessarily imply increased risks. It is possible that smaller AI models with similar capabilities could also be used in harmful ways. Such an approach is also not future-proof, given the way in which technology evolves – to be sure, it is unlikely to keep up with innovation in processor architecture and training methodologies.

In seeking to develop further thresholds to categorize a dual-use foundation model, we encourage the Commerce Department, through the National Institute of Standards and Technology, to work to establish criteria based on the capabilities of a model.