March 27, 2024

Mr. Bertram Lee
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Avenue NW
Washington, DC 20230

Re: Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights

Dear Mr. Lee,

On behalf of the Center for Data Innovation (datainnovation.org), I am pleased to submit this response to the National Telecommunications and Information Administration's (NTIA) request for comment on the potential benefits, risks, and implications of dual-use foundation models for which the model weights are widely available, as well as policy and regulatory recommendations pertaining to those models.[1]

The Center for Data Innovation studies the intersection of data, technology, and public policy, and formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, artificial intelligence, and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

In this submission, we encourage U.S. policymakers to learn lessons from past debates about dual-use technologies, such as encryption, and refrain from imposing restrictions on foundation models with widely available model weights (i.e. "open models") because such policies would not only be ultimately ineffective at addressing risk, but they would slow innovation, reduce competition, and decrease U.S. competitiveness. NTIA should use an evidence-based approach to addressing AI risks and avoid broad rules that would negatively impact the ability to develop open models. Moreover, U.S. policymakers should defend open AI models at the international level as part of its continued embrace of the global free flow of data.

Sincerely,

Daniel Castro
Vice President, Information Technology and Innovation Foundation (ITIF)
Director, ITIF's Center for Data Innovation

[1] "Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights," Federal Register, February 26, 2024, https://www.federalregister.gov/documents/2024/02/26/2024-03763/dual-use-foundation-artificial-intelligence-models-with-widely-available-model-weights.

1. HOW SHOULD NTIA DEFINE "OPEN" OR "WIDELY AVAILABLE" WHEN THINKING ABOUT FOUNDATION MODELS AND MODEL WEIGHTS?

"Open" should refer to models and model weights that are made freely available, for any purpose, for anyone. In other words, anyone may use, distribute, customize, or improve the model and model weights at no cost. Some models and model weights may be partially open, meaning that they have some restrictions. For example, a developer may release a model under license terms that only allow for non-commercial use or require that any derivative works retain the same license. These restrictions may also prohibit certain harmful uses. For example, Llama 2's acceptable use policy explicitly prohibits using the model for "illegal or unlawful activity or content," "activities that present a risk of death or bodily harm to individuals," and to "intentionally deceive or mislead others."[2]

"Widely available" should refer to models and model weights that are freely accessible on the Internet. This definition mirrors how the IRS uses the term.[3] According to IRS regulations, a document is "widely available" when an organization 1) informs users that the document is available and provides download instructions; 2) provides the document in a format that allows exact reproduction of the original; and 3) allows any user to access the document without special hardware or software and without payment of any fee.[4] NTIA should incorporate similar definitions in its work.

NTIA should not base the widely available definition on the number of downloads or the number of entities making the model or model weights available for download. Instead, as in the IRS definition, it should be based on meeting certain specific criteria related to its potential discoverability and accessibility on the Internet. For example, NTIA should consider a model to be widely available if a developer makes its model available on its own website or publishes it to a third-party website, such as a publicly accessible GitHub or HuggingFace repository.

Some repositories like HuggingFace allow developers to provide gated models, meaning that users must request permission before they can access them. In some cases, the developer can authorize the platform to automatically approve these requests. In other cases, the developer may review requests and only grant them to certain individuals. NTIA should consider gated models widely available if they have automatic approval.

---

[2] "Acceptable Use Policy," n.d., Meta, https://ai.meta.com/llama/use-policy/.
[3] "Public Disclosure and Availability of Exempt Organizations Returns and Applications: Exemption Where Organization Makes Documents 'Widely Available'," IRS, December 4, 2023, https://www.irs.gov/charities-non-profits/public-disclosure-and-availability-of-exempt-organizations-returns-and-applications-exemption-where-organization-makes-documents-widely-available.
[4] "26 CFR § 301.6104(d)-2 - Making applications and returns widely available," Cornell Law School, n.d., https://www.law.cornell.edu/cfr/text/26/301.6104(d)-2.

2. HOW DO THE RISKS ASSOCIATED WITH MAKING MODEL WEIGHTS WIDELY AVAILABLE COMPARE TO THE RISKS ASSOCIATED WITH NON-PUBLIC MODEL WEIGHTS?

NTIA's focus is on foundation models with widely accessible weights that pose significant risks. However, determining whether a model presents such risks can be challenging. While some risks may be apparent from the outset, many are not, and even for those risks that are apparent, it may not be clear if they are significant. There is no standardized benchmark for assessing whether a specific AI model poses a significant risk. Indeed, opinions vary widely on what constitutes a serious threat. For instance, a terrorist constructing a bomb unquestionably poses a national security risk. By this definition, a large language model (LLM) offering guidance on bomb-making falls into this category. Yet, similar information is readily available on the Internet. Treating a search engine that can uncover this data differently from an LLM that does the same would be inconsistent and ineffective policy.

Therefore, instead of focusing solely on the absolute level of risk, NTIA should consider the relative increase in risk resulting from open AI models. These risks exist along a spectrum. On one end are scenarios where open AI models introduce entirely new risks that would not exist without them. On the other end, open models merely reduce barriers to existing risks. However, many instances exist where open AI models merely lower barriers to risk in relatively inconsequential ways. Thus, NTIA should prioritize oversight and research on cases where open AI models significantly enhance performance in tasks posing serious security risks compared to non-open AI models.

Additionally, NTIA should recognize that even when AI models introduce new risks, the response is not necessarily to restrict the technology. Consider the invention of the automobile: crime existed before personal vehicles became widespread, but they also facilitated criminal activities, as exemplified by John Dillinger's infamous use of getaway cars in bank robberies. However, the government's response was not to ban vehicles but rather to rethink approaches to crime prevention, leading to the establishment of the FBI and the allocation of new resources to law enforcement.

3. WHAT ARE THE BENEFITS OF FOUNDATION MODELS WITH MODEL WEIGHTS THAT ARE WIDELY AVAILABLE AS COMPARED TO FULLY CLOSED MODELS?

Foundation models with model weights that are widely available, i.e., open models, provide an important pathway for innovation since anyone can freely use and modify the model for research and commercial purposes. This collaborative process accelerates development, cuts costs, and democratizes access to AI technology. For example, developers can fine-tune open models for specific applications thereby facilitating adoption. In some cases, commercial providers of closed models may be unwilling or unable to address the AI needs of a specific sector or community, but if members of that sector or community have access to open models, they can customize them for their own purposes. For example, a non-profit organization may adapt an AI model for a specific healthcare application or educational use case.

Open models not only enable research and development on more capable AI models, but they also allow progress on other goals such as explainability, bias, safety, and efficiency. Here again having open models is useful because commercial providers cannot always focus on improving every useful feature. Allowing others to build on open models allows developers and researchers focused on specific improvements to contribute to advancements in the field, including critical ones like security and safety. These advancements can benefit both open and closed models, as developers making closed models can still integrate the ideas and techniques developed for open models.

Open models also provide opportunities for those learning about AI to better understand how these systems work by being able to "look under the hood" and experiment in ways that are not usually possible with fully closed models. This type of accessibility means more people can understand exactly how AI works, which helps to demystify the technology and increase public acceptance. For example, concerns about fairness and bias could impede adoption of AI in fields such as criminal justice or public administration, but using open models could mitigate some concerns because anyone would have the opportunity to conduct their own independent third-party testing. In addition, having access to not just open models, but open-source code and training data, allows for greater understanding of AI by students and practitioners in the field, as well as increased opportunities for customization and collaboration.

Competition between open and closed models provides more user choice. History shows that there is a market for both open and closed approaches, such as in operating systems, software, and mobile device ecosystems, for both consumers and enterprises. Competition between open and closed models encourages innovation that benefits users. Policymakers should provide a level playing field, and neither penalize nor favor open models over closed models.

There are open foundation models today, and there will almost certainly be open foundation models in the future. The question is whether these models will be created and led by U.S. developers and reflect U.S. values.[5] Imposing restrictions on open models for U.S. firms and developers will hurt U.S. competitiveness as other nations, including China, would fill that void and U.S.-based firms interested in developing open AI would likely consider relocating abroad. Moreover, if foreign firms are creating the state-of-the-art open foundation models, the U.S. government will have less influence over its development and oversight over potential risks.

[5] "A Global Declaration on Free and Open AI," Information Technology and Innovation Foundation, September 13, 2023, https://itif.org/publications/2023/09/13/global-declaration-on-free-and-open-ai/.

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?

The challenges of dual-use, open-source technology are not new. For example, in the 1990s there was widespread debate over how policymakers should address the risks of strong encryption.[6] At the time, U.S. policy restricted the export of advanced encryption technology. In 1991, Phil Zimmermann created Pretty Good Privacy (PGP), an open-source public-key encryption program to enable anyone to sign and encrypt files and messages. After Zimmerman published his software, the U.S. government investigated him for violating export controls (PGP used a minimum of 128-bit keys, enabling significantly stronger encryption than was allowed to be exported at the time), but eventually dropped its case. Policymakers came to realize that restricting encryption technology was both impractical (because limiting its lawful distribution had no impact on bad actors getting access to the technology) and counterproductive (because it would limit U.S. users from using encryption to secure their data and communications and hurt U.S. businesses that could not integrate advanced encryption technology into their products and services).

Policymakers face a similar scenario today with dual-use foundation models. Just as there were some people pushing for restrictions on encryption in the 1990s, primarily in the interest of national security, so too are there some today advocating for similar restrictions on AI models, especially open AI models, for similar reasons. While bad actors might make use of this technology for harmful activities, there are almost always better ways to mitigate these risks than restricting AI models, especially when such restrictions would negatively impact many beneficial uses of the technology. Therefore, rather than focusing on general restrictions aimed at foundation models for a broad set of risks, policymakers should instead focus on specific countermeasures aimed at specific risks. In many cases, the most effective countermeasures may not be related to imposing restrictions on AI models, but instead addressing the problem somewhere closer to the problematic behavior. Moreover, U.S. policymakers should recognize that there is little they can do to stop open models developed and released abroad.

The research by the National Institute of Standards and Technology (NIST) on AI test, evaluation, validation, and verification (TEVV) will support AI safety in both closed and open AI models. As part of this effort, NTIA should direct the NIST AI Safety Institute to include a workstream on assessing AI memorization—or the condition where AI models output training data verbatim. In many cases, AI memorization is a feature, such as to ensure that a model produces accurate, factual output. However, in other cases, AI memorization can be a problem, such as if it reveals private information from training data. Developing protocols for efficiently and effectively assessing memorization in AI models will help developers identify and address potential risks.

[6] Daniel Castro, "Why New Calls to Subvert Commercial Encryption Are Unjustified," July 13, 2020, Information Technology and Innovation Foundation, https://itif.org/publications/2020/07/13/why-new-calls-subvert-commercial-encryption-are-unjustified/.

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?

The U.S. government has historically supported the global free flow of data because cross-border data flows are essential for the digital economy. Restricting data flows, such as with data localization requirements, can slow innovation, raise costs, and hurt economic growth. U.S. policymakers should continue to uphold this position for AI, specifically by opposing restrictions on the global free flow of AI model weights (which are a form of data). For example, cross-border restrictions on widely available model weights could limit access to AI models developed and used by U.S. companies thereby restricting market access for U.S. firms. In addition, some governments could impose restrictions on sharing AI model weights to limit free speech.

The U.S. government should also oppose regulations that focus on the development and sharing of AI models, including those with widely available weights, rather than the use of AI models. Imposing restrictions on the development and sharing of AI models can make open-source development impractical, especially when there is no central developer managing the project. Instead, policymakers should focus any regulations on specific high-risk uses of AI, so that any rules and requirements relate to the actual context in which an entity uses the AI system. For example, it would be better for the Department of Transportation to create specific regulations about the safety of autonomous vehicles rather than have broad regulations about the development of AI-based image recognition systems that have thousands of different potential uses.

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?

The Department of Commerce should create evidence-based policy. In this case, there is no evidence that AI models that exceed certain computational thresholds or require a certain amount of computational resources present novel risks that require extraordinary treatment. In addition, there is new evidence that the risk of sudden emergent abilities in AI models is less likely than previous thought and that better evaluation can help assess the capabilities of AI models.[7]

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